Submitted to the Department of Biology
in partial fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Biology

at the
Massachusetts Institute of Technology

June 2011



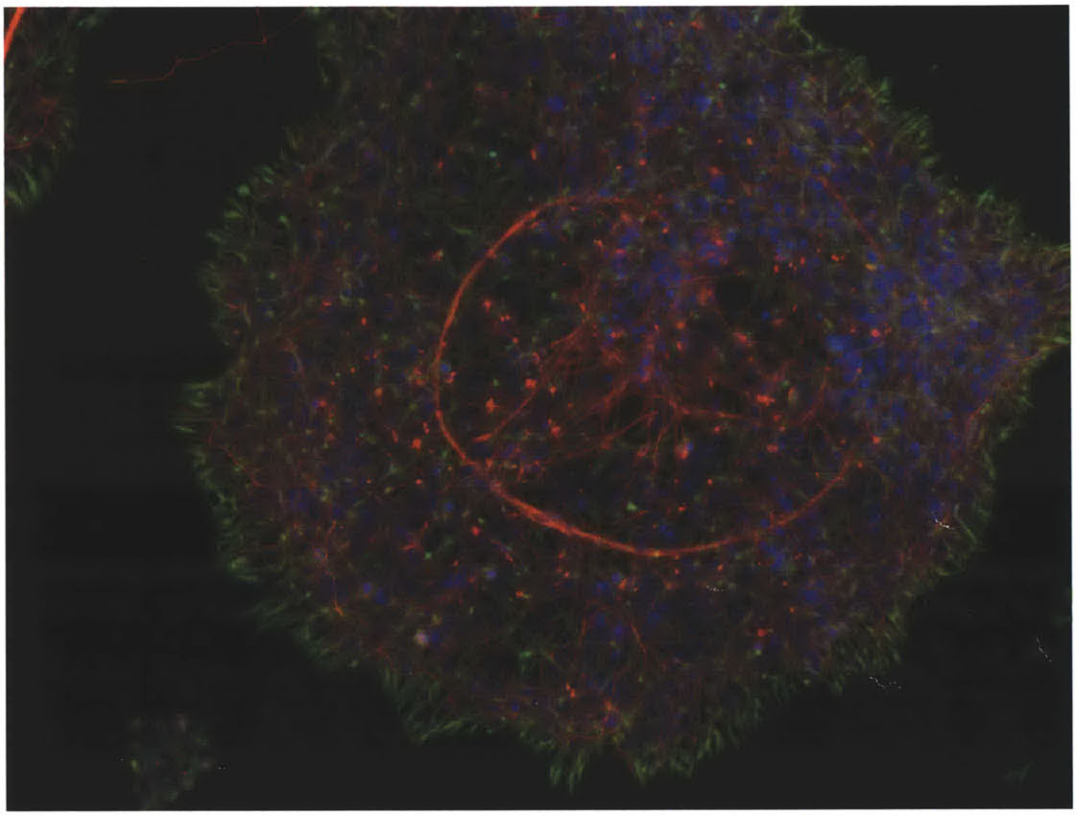


Signature of Author: _____
Department of Biology
June 2011

Certified by: _____
Rudolf Jaenisch
Department of Biology
isor

Accepted by: _____
Alan D. Grossman
Co-Chair, Department Committee on Graduate Students

# INDUCED PLURIPOTENCY AND EPIGENETIC REPROGRAMMING BY DEFINED FACTORS



**THESIS PROPOSAL**

Bryce W. Carey
B.S., Cell Biology
University of California Davis, 2003

# INDUCED PLURIPOTENCY AND REPROGRAMMING BY DEFINED FACTORS

by

Bryce W. Carey

**Abstract:** The process by which the totipotent zygote undergoes development into an adult organism using a single genome is the foundation for epigenetics and cellular differentiation. Somatic cell nuclear transfer experiments (SCNT) provided unequivocal proof of nuclear equivalency between adult somatic cells. However the molecular mechanisms of somatic cell reprogramming have remained largely a mystery. Recent advancements in epigenetic reprogramming by defined factors provide new opportunities to explore factors that regulate induction of pluripotency in somatic cells. Nuclear reprogramming by SCNT occurs in an 'indirect' manner by unidentified components within oocyte cytoplasm and requires the destruction of embryos. The introduction of induced pluripotent stem cells (iPS cells) and 'direct' reprogramming methods created a tractable system to both study of the process in vitro and potentially derive personalized pluripotent stem cells free of the practical and ethical concerns surrounding embryonic stem (ES) cells and SCNT. Herein we study mouse somatic cell reprogramming by defined factors and develop novel tools to compare the induced pluripotent state to the gold standard of pluripotency, ES cells. First, we designed reprogramming vectors that minimize the number of viruses required to generate iPS cells, yielding pluripotent cells with minimal genomic alterations from reprogramming factors. This allowed the creation of transgenic "reprogrammable mouse" strains after gene targeting in ES cells, providing a renewable source of somatic cells that can be induced to pluripotency by addition of a drug. In addition we can easily introduce or mate these strains to study unique genetic variants during reprogramming. Third, we study factors that influence the induced pluripotent state, specifically how to generate pluripotent cells with all properties of embryonic stem cells including derivation of "all iPS cell mice" by tetraploid complementation assays. In contrast to previous reports, we find the majority (~ 80%) of iPS cell lines derived from adult somatic cells of varying organs contain the developmental potential of ES cells. This outcome correlated with high expression of Oct4 and Klf4 and low expression of Sox2 and c-Myc during reprogramming. In addition we report that adult mice derived from iPS cells are healthy and do not develop tumors. Together these results suggest in vitro reprogramming to pluripotency by defined factors holds great promise for regenerative medicine.

Thesis supervisor: Dr. Rudolf Jaenisch
Title: Professor of Biology

**Dedication**: I would like to dedicate this to my father James R. Carey for his endless encouragement and support of my pursuits towards a life conducting scientific research. He endures as a model for me both personally and academically as someone who lives with great integrity, absolute honesty and strength of character.

**Acknowledgements:**

I would like to first thank Rudolf for taking me into his lab. It has been a remarkable experience I will never forget. He and all lab members were extremely welcoming upon arrival and provided me with a fantastic, highly supportive environment to conduct my thesis research. Although conducted in a super charged, internationally competitive time surrounding the discovery of induced pluripotent stem cells, Rudolf always maintained his patience, support, and encouragement no matter what was going on around us. His steadfast leadership and long-term commitment ensured the success of all of my projects.

I also have had the privilege of working alongside a good friend and colleague Jacob Hanna, who I would like to thank. His keen insights and vision on many projects kept me from veering off course numerous times. His focus and determination to do the right experiment were an inspiration throughout my time in the Jaenisch lab. These experiments would also not have been possible without several other close colleagues who have also become good friends. I would not have got off the ground without the help of Marius Wernig. His efforts to establish reprogramming technologies in the Jaenisch lab propelled many careers forward including my own, teaching me my first reprogramming experiment and helping me design my first reprogramming vectors. I would like to thank Stella Markoulaki who I worked with on every project to generate mice. She remains a strong supporter and someone who I can always rely on. Caroline Beard provided invaluable help in the beginning stages of my projects; her keen advice and unparalled knowledge of molecular biology will always be remembered as an invaluable asset to the Jaenisch lab. Kibibi Gantz is also someone who helped me in the mouse-generating department. I appreciate her dedication and feel she is a great new addition to the Jaenisch Lab. I would also like to thank Maya Mitalipova who introduced me to human stem cell work. She provided crucial support in the early years of my project. In addition I held close collaborations with Menno Creyghton and Albert Cheng who taught me genome-wide mapping technologies, always showing patience and dedication to our questions. Along the way I was supported experimentally or stimulated intellectually from many post-docs and grad students such as Frank Soldner, Chris Lengner, Kris Saha, JP Kim, Mike Lodato, Leah Medieros, John Cassady, Dina Faddah, Grant Welstead and always received support in some form or another by anyone who I asked for help. This is something that is often taken for granted in the Jaenisch lab!!

Most importantly I want to thank the person who kept me sane throughout this process, my fiancé Leah Cohen. She has been my most steadfast supporter at all times and continually helped me stay balanced outside of lab. I feel so very lucky to have her in my life.

Finally I want to thank my thesis committee members Tyler Jacks, Peter Reddien and Laurie Boyer. They always encouraged me to keep at it and to be patient so I can do the right experiment. Their skepticism also helped me stay on course at critical junctures. In addition I want to thank Alex Meissner for returning to the Whitehead to generously serve as my outside committee member.

## Table of Contents

## LIST OF FIGURES

### CHAPTER 1



### CHAPTER 2



### CHAPTER 3



### CHAPTER 4

# Chapter 1: Introduction

In vitro somatic cell reprogramming has taken on new meaning in the last few years with the introduction of induced pluripotent stem cells (iPSCs). Remarkably, the human genome contains over 22,000 protein-coding genes and yet just three or four are sufficient to reprogram adult somatic cells into an embyronic-like state. These findings raised considerable interest in the nature and stability of cellular differentiation, as had been highlighted by the pioneers of reprogramming research using somatic cell nuclear transfer (SCNT). The reprogramming of cell identity can now be observed and studied in a petri dish, providing a model system to test the rules required for rewiring of both gene expression and chromatin in order for cells to take on new functions. In this chapter I will describe two main areas of scientific research that provide the conceptual and biological foundation of my thesis research. I will summarize what is understood to date about nuclear reprogramming and the establishment, propagation and maintenance of pluripotency in vitro. Towards this goal I will end by discussing similarities and differences between pluripotent cells created by current reprogramming technologies and speculate as to whether there is sufficient evidence to suggest defined factors re-establish cell identities with molecular and developmental potential similar to what has been described by isolation of pluripotent cells from the embryo.

## 1.1 Reprogramming by nuclear transfer

Upon fertilization the single cell zygote proceeds through a defined set of cellular divisions using one genome to guide and instruct the development of an adult organism containing over 200 diverse cell types. The process of cellular differentiation is unidirectional and fundamentally epigenetic in nature. As differentiation ensues there is a

concomitant reduction in cellular potency, defined herein as the sum of developmental options accessible to the cell. Totipotency, is the ability to form all lineages of an organism, and in mammals only the zygote and the first cleavage blastomeres contain this potential. Cleavage divisions allow the zygotic genome to become activated, and as this period ceases, compaction begins leading to formation of the blastocyst. It is at this stage that the first segregation of developmental lineages is accomplished. The outer layer of cells form an epithelium, the trophectoderm, responsible for generating the extraembryonic tissue, and the interior epiblast cells of the inner cell mass (ICM), develops into the embryo proper. Pluripotency, or the ability to give rise to all somatic cells including the germ lineage of an adult animal, exists only transiently during development but has been experimentally demonstrated to exist within the epiblast cells of the ICM [1, 2]. This relatively unspecified cell state can be immortalized in vitro in the form of embryonic stem cells (ESCs) derived from explantation of the inner cell mass (ICM) of preimplantation blastocyst embryos, a pluripotent cell type that will be discussed in later sections [3-5]. Emerging from a brief period of pluripotency the postimplantation embryo begins gastrulation generating tissues and organs with highly specialized functions. Some tissues, such as the hematopoetic, contain adult stem cells that are considered multipotent, giving rise to differentiated cells only within their lineage but not others. The vast majorities of cells however, become terminally differentiated, rendered unipotent, and only give rise to identical daughter cells throughout their lifetime.

Beginning more than a half-century ago, scientists tested whether a reduction in developmental potential during differentiation is an irreversible process, and applied the technique of somatic cell nuclear transfer (SCNT) to test whether the capacity for embryonic development still existed within a somatic cell nucleus following differentia-

tion. These experiments were some of the first attempts at nuclear reprogramming, which for the purposes of this thesis will be defined as a *gain* in developmental potency. Somatic cell nuclear transfer (SCNT) describes the technique by which a somatic nucleus is transferred to an enucleated oocyte to generate cloned embryos or animals achieved by reactivation of developmental gene expression programs without altering the underlying DNA sequence. The successful derivation of cloned amphibians from *Xenopus* intestinal epithelial cell nuclei argued somatic cells contain all basic genetic information required for development and revealed that epigenetic marks acquired during differentiation are fully reversible, providing compelling evidence against the hypothesis that differentiation permanently inactivated or discarded much of the nuclear genome [6,7]. The concept of nuclear equivalency between somatic and embryonic cells emerged. The genomic nucleotide content between these cells is virtually identical with embryonic gene expression programs silenced in somatic cells being capable of reactivation simply by exposure to egg cytoplasm. Only decades later were the pioneering nuclear transfer efforts of Briggs and King along with the foundational cloning experiments of John Gurdon, successfully translated to mammalian systems by McGrath and Solter. Using Sendai-virus-mediated cell fusion, they demonstrated murine zygotic pronuclei could reconstitute an enucleated zygote to generate blastocysts and following embryo transfer, live mice [8]. However following a similar protocol, unsuccessful experiments using nuclei from more developed embryonic cells, such as cleavage stage blastomeres or inner cell mass (ICM), raised concerns as to the ability of mammalian systems to successfully generate cloned adult animals from adult somatic cells [9].

These conclusions were soon proved premature, and rested largely upon the activation status of the egg, as cytoplasm of unactivated oocytes (Metaphase II arrested)

supported reprogramming after nuclear transfer. By electrical-mediated membrane fusion, embryonic or adult ovine cells introduced into unactivated oocytes generated adult sheep [10, 11]. One hypothesis for the failure of previous efforts was that reprogramming factors remained in the cytoplasm in unactivated oocytes, but were sequested upon entry into the cell cycle in the fertilized egg. Data by Wakayama and colleagues support this hypothesis and show in vitro development is markedly reduced in cloned embryos using oocyte cytoplasm transitioning from meiosis to G1 [12]. Interestingly, researchers recently addressed whether entry into the cell cycle sequesters reprogramming activities that would otherwise be present within the egg cytoplasm of unactivated oocytes. Eggan and colleagues reversibly arrested activated mouse zygotes using a microtubule depolymerizing agent nicodozole in combination with a proteasome inhibitor MG-132 to delay the metaphase-to-anaphase transition allowing chromosome capture and replacement [13]. Using either embryonic or somatic cell chromosomes, also arrested in mitosis, ~ 60% of embryos that received new chromosomes entered the cell cycle and initiated further cleavage to blastocyst stage embryos with ~ 30% transferred to pseudo pregnant females producing adult mice [13]. Shortly thereafter this protocol was applied to mouse blastomeres [14]. These data argue that reprogramming factors are removed following zygotic activation, most likely by nuclear condensation, but remain and accessible even after the initiation of cleavage divisions.

Although many animals have now been produced following transfer of somatic nuclei into enucleated oocytes: sheep [11], cows [15], pigs [16, 17], goats [18], and mice [12], there was much debate over the identity of donor cells, and whether successes were largely due to rare stem cells present within the culture. Intriguing evidence also came from studies comparing nuclear transfer of unique donor cells in mouse. Donor cells with more embryonic origins, such as blastomeres or embryonic stem cells, show a marked increase in

generating cloned mice (5-25% of blastocysts transfered) when compared to cumulus cell or tail-tip fibroblast nuclear donor embryos (1-3% and 0.5% respectively) [12, 19-23]. These data suggest the embryonic genome may be inherently more amenable to epigenetic reprogramming than somatic cells. One hypothesis is that embryonic genes essential for development, such as Oct4 [24] are already expressed in embryonic cells. Therefore faulty or incomplete reactivation was less likely to occur compared to somatic cells, whose embryonic genes are epigenetically silenced and must be reactivated to an embryonic-like epigenetic state. Evidence in support of this hypothesis has been observed when comparing gene expression profiles between ESC- or somatic cell derived cloned embryos [25].

To definitively address the hypothesis that only rare populations of somatic cells were amenable to cloning, and that highly specialized somatic cells may actually prove *intractable* to cloning, much effort was put forth to generate cloned animals from terminally differentiated cells that could be retrospectively identified by permanent alternations to the genome. B and T lymphocytes proved suitable for this type of analysis as heavy and or light chain rearrangements to the B- or T-cell receptors yielded cells with permanent genetic identification reflecting their stage of differentiation reached [26]. Similar approaches using the genetic marking of terminally differentiated olfactory neurons demonstrated even cells that have exited the cell cycle can re-enter development [27]. The success of these studies required modifications to existing protocols of murine SCNT, using a two-step procedure that began first by generating ES cells from cloned blastocysts, and then was followed by the generation of adult monoclonal mice using the tetraploid embryo complementation method.

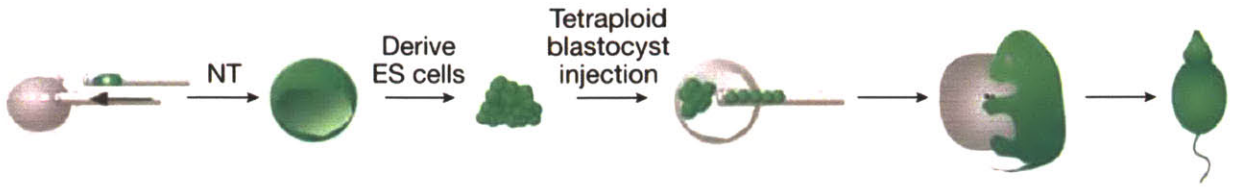



**FIGURE 1: Two-step derivation of adult mice from terminally differentiated cell nuclei** (Eggan et al., 2004)

The NT-ES cell lines and the derived monoclonal mice carried rearrangements of the IgH and IgL immunoglobulin loci or TCRα and TCRβ chain rearrangements, demonstrating their origin from a mature B or T cell, respectively [26]. Later experiments used different adult, genetically modified, lymphocytes and achieved cloned mice using a one-step protocol [28].

The reprogramming activities of the oocyte or early embryonic cells capable of reprogramming such a diverse set of somatic cells sparked intense interest in the molecular identity of reprogramming factors. Following the discoveries and technological advances described previously, and along with the isolation of human embryonic stem cells in 1998, cloning captured the imagination of many who believed that "personalized" genetically identical pluripotent cells could be generated to usher in an unprecedented era of regenerative medicine [5, 29]. Alternative reprogramming strategies other than nuclear transfer will be discussed in detail after a review of pluripotency and embryonic stem cells.

### 1.2: Signaling pathways to stabilize pluripotency in vitro

Embryonic stem cells (ESCs) self-renew indefinitely and with each division, create two daughter cells with identical potential, all while retaining the ability to generate somatic cells of all three germ layers both in vitro and in vivo. ESCs colonize the ICM following microinjection of blastocysts, and after embryo transfer to pseudo pregnant mothers produce functional soma and germ cells [30]. The initial derivation of ESCs from normal blastocyst stage embryos was performed using the "129" mouse strain [3,4]. Isolation and maintenance of pluripotent ESCs was first performed using media containing fetal bovine serum and a layer of mouse embryonic fibroblast (MEFs) feeders. Feeders were later found to secrete an essential cytokine, LIF, whose signaling through the Jak-Stat3 pathway is necessary for normal ESC self-renewal [31]. LIF binding to the LIF receptor causes heterodimerization with gp130, which results in the activation of the receptor-associated tyrosine kinase Jak [32]. Activated Jak phosphorylates several tyrosines of gp130, which serve as docking sites for Stat family transcription factors. In the absence of LIF, Stat3 activation is sufficient for maintenance of undifferentiated ESCs [33]. Interestingly, mutant embryos deficient in the LIF gene form normal ICM [34], supporting the observations that although some requirements for pluripotency in vitro are similar, some aspects remain distinct from what is necessary for growth and expansion of the epiblast *in vivo*.

Serum factors necessary for ESC growth have been identified as bone morphogenic factor 4 (BMP4), whose downstream signaling through Smad family of transcription factors induces inhibitor of differentiation (Id) genes [35]. This pathway was found to maintain ESC self-renewal by inhibiting lineage-specific transcription factors, which in turn, enable the self-renewal response to LIF/Stat3 [35]. Importantly, without LIF, BMP acts as a differentiation promoter and generates non-neural epithelial-like cells [36,37], suggesting the integration of these pathways reinforce self-renewal of ESCs. Austin

Smith and colleagues hypothesized a "ground state" of pluripotency could be isolated by suppression or neutralization of differentiation signals generally, which would insulate the ESC transcriptional network to allow self-renewal. To this end, investigation into the signaling pathways sufficient for ESC maintenance turned to previously characterized pathways that had been reported to induce differentiation of ESCs [36, 38].

A key observation was that autoinductive signaling by fibroblast growth factor-4 (FGF4) and the mitogen-activated protein kinase pathway (via ERK1/2) induces differentiation of ESCs, which precluded long-term growth of undifferentiated ESCs in defined (serum-free) media. Neither BMB4 nor LIF signaling inhibits activation of ERK1/2 in ESCs, and importantly, inhibition of ERK1/2 had previously been reported to enhance the maintenance of ESCs [38]. Moreover, genetic impairment or selective chemical blockade of this pathway can sustain self-renewal of mouse ESCs in the absence of LIF signaling [39]. In a landmark study by Smith and colleagues they described the requirements for maintenance of what was termed the "ground state" of pluripotency using defined media with small molecule inhibitors (termed "2i") of both protein kinases ERK1/2 and GSK3β, which when combined with LIF, supported maintenance of fully pluripotent ESCs without serum or feeders [40] (**Figure 2**). The inhibition of GSK3β was proposed as a means to increase the biosynthetic capacity of ESCs as GSK3β has been identified as a central inhibitor of various anabolic processes in the cell. Moreover inhibition of GSK3β mimics activation of Wnt signaling, which has also been shown to promote pluripotency [41-44]. An alternative to "2i + LIF" conditions was also described, termed "3i,"which also includes an inhibitor of the FGF receptor, and does not require addition of LIF for propagation of pluripotent ESCs [40] . Suppression of these pathways

during development of mouse blastocysts indicate development of the ICM is not impeded and that derivation of ESCs from single epiblast cells is greatly enhanced [45].

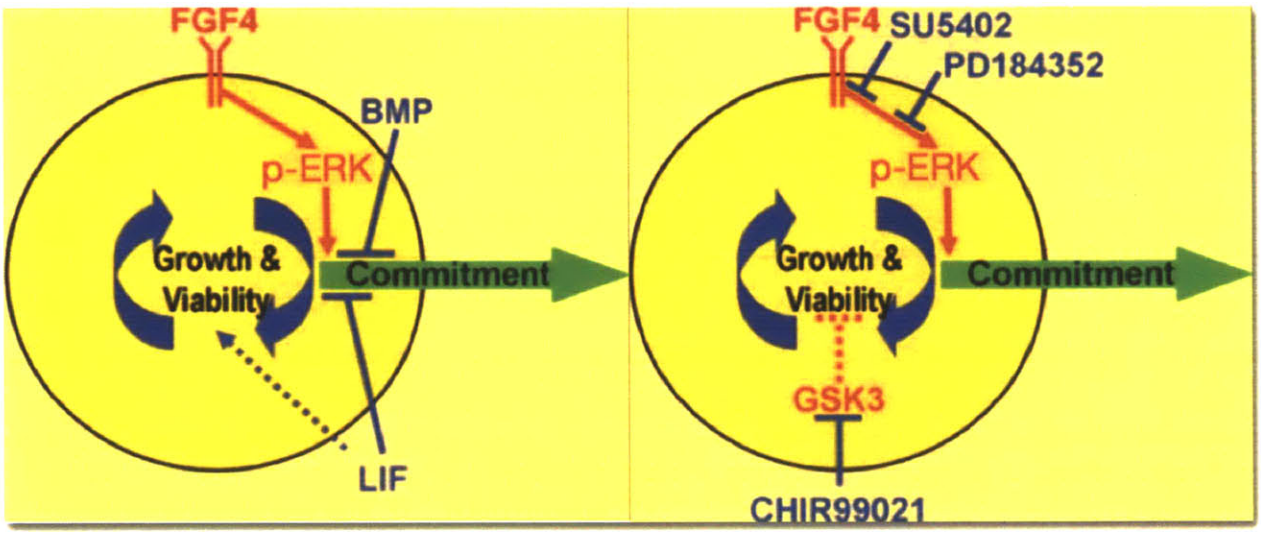



**FIGURE 2: Signaling principles of the "ground state" of pluripotency** Self-replication of the pluripotent state can occur when inductive phospho-ERK signaling is counteracted downstream by LIF and BMP. In chemically defined conditions, antagonists of this pathway augment self-renewal when phospho-ERK (p-ERK) is suppressed by maintaining cellular growth capacity and additionally suppressing differentiation (Ying et al., Nature 2008).

The signaling principles of ground state of pluripotency have broader applications beyond permissive murine strains. Application of the empirically established ESC derivation protocol to inbred mouse strains other than 129 or C57BL/6, for example the non-obese diabetic (NOD) strain or even other rodents such as rat, do not yield pluripotent cells similar to ICM-derived ESCs. However under "3i" or "2i+LIF" conditions rat or NOD mouse E3.5-E4.5 blastocysts give rise to ESCs with molecular and developmental character reflecting authentic 129 mouse ICM-derived ESCs, such as signaling de-

pendence on LIF/Stat3 and formation of adult chimeras with contribution to the germline [46-49]. Because ESCs could be isolated from previously refractory strains and species in vitro, these data argue instability of the pluripotent state in vitro can be overcome and stabilized with appropriate signaling stimulation [50]. These data also highlight the importance of endogenous genetic determinants, which can be lacking in certain genetic backgrounds.

**1.3: Core transcriptional circuitry of the pluripotent state**

The previous section concentrated on the signaling pathways that enable the establishment and maintenance of the pluripotent state of ESCs in vitro. These pathways collectively act to stabilize the global transcriptional program of ESCs, which is beginning to be understood. The identification of key transcription factors critical to establishing pluripotency in vitro and in vivo is the focus of this section. Classical genetic analysis has been utilized to provide the most compelling evidence to date for a role in pluripotency, therefore I chose to focus on several genes that were identified or whose function was validated *in vivo* as a critical determinant of establishment or specification of pluripotent cells of the embryo. Following a brief introduction of each factor, I will summarize more recent work characterizing their role in integrating and maintaining the ES cell transcriptional network.

*Oct4*: Expression of this POU-domain transcription factor is restricted to oocytes, blastomeres, pluripotent early embryo cells, and the germ cell lineage [51-54]. Null Oct4 embryos die in utero during the peri-implantation stages [24] and in vitro cultures of the ICM demonstrate it is responsible for the establishment of pluripotent cells [24, 55]. Loss of Oct4 generates only cells of trophoblast origin in vitro, suggesting it may actively inhibit differentiation towards this lineage. Importantly, as a transcription factor, it can

act to either repress or to activate target gene transcription along with partner factors [56-60].

*Sox2:* Sox (SRY-related HMG-box) family proteins are transcriptional regulators that typically rely on association with partner factors to affect target genes. The expression pattern of Sox2 is similar to Oct4, marking the ICM, epiblast, and germ cells, however, Sox2 is also expressed in non-pluripotent tissues such as the multipotential cells of the extraembryonic ectoderm and uncommitted dividing stem and precursor cells of the developing central nervous system (CNS)[61] . Sox2 null embryos die at the time of implantation due to a failure of epiblast (primitive ectoderm) development [62]. ESCs deleted for Sox2 results in trophectoderm and primitive endoderm-like cells suggesting that like Oct4, it is essential for pluripotency. Sox2 is frequently found at gene promoters as a heterodimer with Oct4 and co-binding has been shown to be necessary for synergistic gene activation at targets such as Fgf4 [63]. Co-binding was also reported at the Sox2 and Oct4 genes themselves [64-66] in addition to another canonical pluripotency marker, Nanog.

*Nanog*: A highly divergent homeodomain-containing protein that is activated upon compaction with expression specifically demarcating the nascent epiblast [67-69]. Two groups initially characterized the function of Nanog in ESCs. Yamanaka and colleagues used in silico differential display of expressed sequence tag (EST) libraries from mouse ES cells and various somatic cells to isolate a set of pluripotency-associated genes. Functional validation demonstrated Nanog-null embryos can develop blastocysts but cannot form a functional epiblast, with explantation of null (-/-) ICM failing to generate ESCs [70]. In a second study Chambers et al. analyzed $10^5$ independent cDNA plasmids for factors capable of rescuing LIF receptor -/- ESCs. These genetically modified ESCs used IL6/sIL6R to activate gp130 and Stat3, and were screened by episomal

transfection for self-renewal in media lacking cytokines [68]. Ectopic expression of Nanog endowed transgenic ESCs with the ability to grow in the absence of LIF-receptor, and in addition, these cells became refractory towards differentiation signals such as exposure to all-*trans* retinoic acid. Interestingly, Nanog expression is maintained through the blastocyst stage in Oct4 null embryos indicating that transcriptional activation of Nanog during this period does not require Oct4 [68].

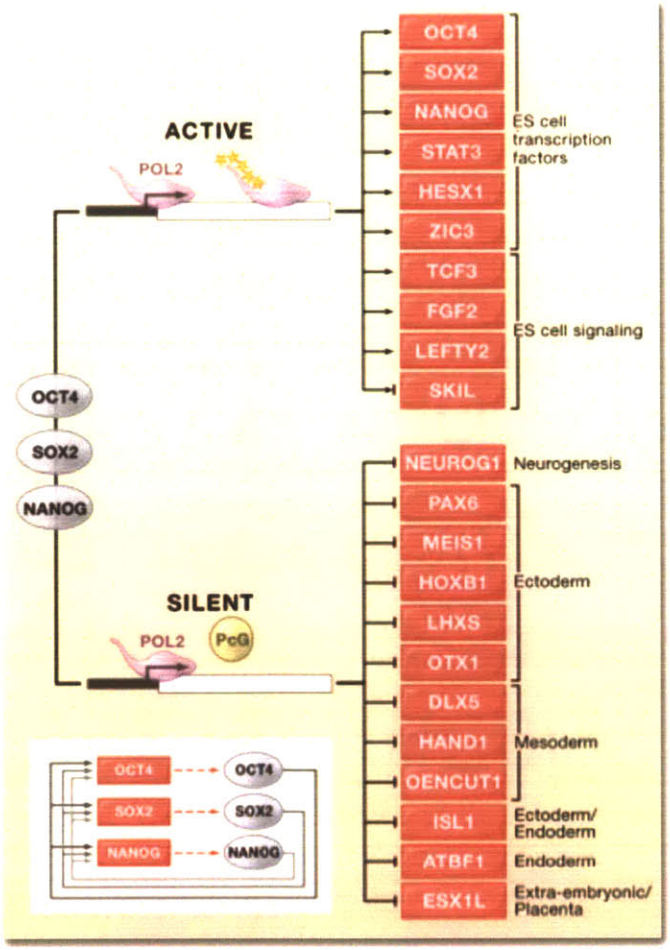
Nanog has a unique role in maintaining ESCs in vitro. It has been described as a "licensing" factor of pluripotency because, surprisingly, established ESCs can be maintained upon conditional deletion of Nanog [71]. Nanog expression fluctuates in ESCs, which allows partitioning into two populations of $Nanog^{low}$ or $Nanog^{high}$, with $Nanog^{low}$ cells showing a higher propensity for differentiation. In general, Nanog null ESCs have a lower threshold for differentiation suggesting it buffers against signals to exit pluripotency [71]. Injection into blastocyst embryos demonstrates Nanog null ESCs are capable of multi-lineage differentiation however fail to contribute to the germ lineage [71].

*Klf4*: Kruppel-like transcription factors (Klfs) are zinc finger proteins that share homology with the Drosophila embryonic pattern regulatory gene *Krüppel* [72]. Although expression of Klf family of transcription factors is not exclusive to pluripotent cells, three Klfs are expressed in ESCs: Klf2, Klf4, and Klf5. In ESCs they most likely harbor redundant functions because loss of pluripotency is only apparent after RNAi-mediated knockdown of all three [73]. Klf5, but not Klf4 or Klf2, knockout embryos show early embryonic leathlity, with null ICM failing to give rise to stable ESCs in vitro [74]. These factors may play a role in helping enforce gene expression upon Oct4 and Sox2 targets. For example Klf4 was reported as an essential factor in the activation of Lefty, a direct Oct4 target, which was unable to be activated with Oct4 and Sox2 alone [75].

Given the importance of these factors in the maintenance of pluripotency a key question is how they act on their transcriptional targets globally. Pioneering studies using genome-wide location analysis (ChIP-ChIP or ChIP-Seq) in human ESCs, initially investigating Oct4, Sox2 and Nanog, has demonstrated pluripotency factors form a core transcriptional regulatory network that governs both pluripotency and the control of differentiation [76, 77]. The results revealed three main observations: First, each factor (Oct4, Sox2, and Nanog) bind the promoters of the other two factors in addition to their own, forming an interconnected, autoregulatory feed-forward loop. Second, all three factors often co-occupy the same sites in the promoters of their target genes. Finally, the list of bound targets can be separated into active or inactive, with the latter being "poised" for activation upon ESC differentiation (**Figure 3**) [76]. Active genes are typically involved in maintenance of ESCs, such as Stat3, Nodal, and a Wnt target Tcf3, however a surprising number of genes bound by Oct4, Sox2 and Nanog are transcriptionally silent. Many of these inactive genes are important transcription factors associated with lineage specification, for example Pax6 (Ectoderm), Sox6 (Mesoderm), and Onecut1 (Endoderm). This suggests that Oct4, Sox2, and Nanog maintain pluripotency not only by activating genes responsible for self-renewal, but also by directly repressing genes associated with differentiation. This hypothesis is supported by genome-wide location analysis of lysine 27 tri-methylation of histone H3 (H3K27me3), catalyzed by the Ezh2 subunit of Polycomb Repressive complex 2 (PRC2), which shows co-enrichment at many silenced developmental regulators co-bound by Oct4, Sox2, and Nanog in ESCs [78, 79]. The genome-wide binding patterns of Oct4, Sox2, or Nanog have been extended to additional pluripotency-associated factors as well as global transcriptional regulators such as p300 and architectural proteins such as CTCF [80, 81]. Both studies suggested there are regions extensively occupied by ESC-specific transcription factors and that actively

expressed genes, as compared to silent or repressed genes, correlated with binding of multiple factors, whereas the unexpressed targets typically have a single pluripotency factor bound in ESCs [81]. Remarkably, the densest binding locus was a previously identified distal enhancer of Oct4 with 11 bound transcription factors [80]. These binding patterns suggest multiple factors reinforce transcriptional activation of the pluripotency network by simultaneously acting on common targets.

**Figure 3: Model of Core ES Cell Regulatory Circutry** The Oct4, Sox2, and Nanog transcription factors (blue) occupy actively transcribed genes, including transcription factors and signaling components necessary to maintain the ES cell state. The three regulators also occupy silent genes encoding transcription factors, that, if expressed, would promote other more differentiated states. At this latter set of genes, RNA pol II (POL2) initiates transcription but does not produce complete transcripts due to the repressive action of PcG proteins [84] Figure below adapted from Boyer et al., *Cell*. 2005.



More precise genetic analysis revealed multiple feedback mechanisms exist between different pluripotency factors. Sox2 conditional deletion in ESCs causes immediate differentiation however this phenotype is rescued by ectopic expression of Oct4, suggesting the role of Sox2 in ESCs may be to maintain expression of Oct4 [82]. Ectopic expression of Nanog is also capable of rescuing RNAi depletion of pluripotency factors such as Esrrb, Tbx3, Dppa4 and Tcl1, whose loss would otherwise induce differentiation, by maintaining expression of Oct4 and Sox2 [83].

Recent studies analyzing the downstream targets of LIF/Stat3 signaling demonstrated the pluripotency transcription factor Tbx3 is capable of maintaining ESCs in the absence of LIF [85]. Interestingly, ectopic expression of Tbx3 induces Nanog but not Sox2 expression, and is sufficient to stabilize Oct4 and maintain undifferentiated ESCs. This work uncovered an apparent bifurcation of the LIF-LIFR-gp130 activation signal with one branch utilizing phosphatidylinositol-3-OH kinase-Akt signaling, instead of Jak kinases, to maintain expression of Tbx3 in ESCs. The other branch of LIF-LIFR-gp130 activation occurs through Jak-Stat3 and primarily leads to induction of Klf4, whose primary downstream target is Sox2 [85] (**Figure 3**).

Using Oct4-inducible null ESCs a close connection to Oct4 and the kruppel-family of transcription factors was uncovered following analysis of acute gene expression changes upon downregulation of Oct4. Two distinct pathways appear to act on each of the three Klfs expressed in ESCs. All (Klf2, -4, -5) are capable, by transgenic expression, of generating LIF-independent ESCs with resistance to differentiation cues such as by BMPs [74, 86]. However only Klf2 is a direct downstream target of Oct4 harboring a consensus octamer binding site proximal to the promoter that is bound by Oct4 in ESCs [86].

Moreover Klf2, but not Klf4 or Klf5, does not require Stat3 for propagation and maintenance of ESCs in the presence of serum, which normally induces differentiation in Stat3 null ESCs. Expression of Klf2 does not change by addition of LIF under "2i" (ERK1/2 & GSK3) conditions, in contrast to Klf4 and Klf5, which are upregulated after just 1 or 4 hr LIF stimulation. These data describe a hierarchy of Oct4 regulated targets with Klf2 being a primary target regulated independent of the LIF/Stat3 pathway, whereas Klf4 and Klf5 are directly downstream of the LIF/Stat3 signal and most likely reinforce the transcriptional output of this key pluripotency pathway [86].

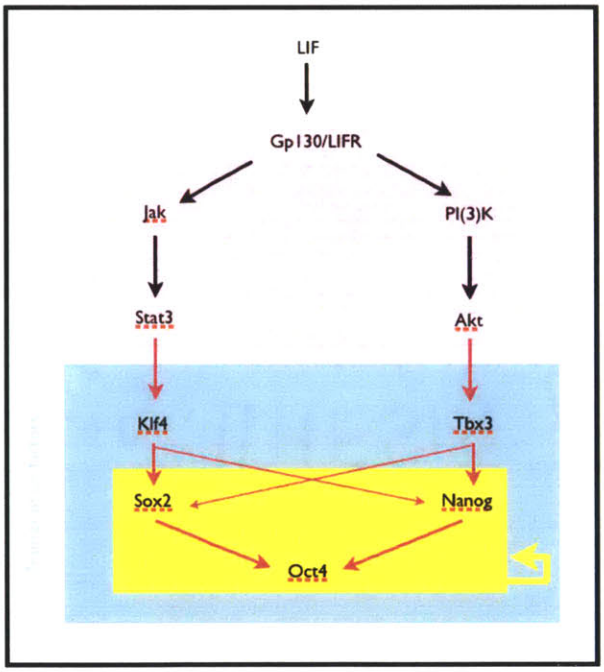



**Figure 4: Parallel circuitry of the LIF signal pathways** Niwa et al. (2009) show the Jak-Stat3 pathway activates Klf4, whereas the PI(3)K-Akt pathway stimulates the transcription of Tbx3 (black). Klf4 and Tbx3 mainly activate Sox2 and Nanog, respectively, and maintain expression of Oct3/4 (red). Transcription of all these transcription factors is positively regulated by Oct3/4, Sox2, and Nanog (yellow).

An additional layer of transcriptional regulation in ESCs is by noncoding RNAs, such as mircoRNAs (miRNA) or large intervening noncoding RNAs (lincRNA). Recent work has incorporated these molecules in the molecular circuitry maintaining pluripotency gene expression programs. An essential role for miRNAs in development was shown by Dicer-deficient mice, which fail to develop [87]. During embryonic development and in established ESCs, subsets of miRNAs are preferentially expressed [88-91]. ES cells genetically deficient in miRNA processing enzymes have defects in differentiation and self-renewal [92, 93]. Investigation into the Dicer null (Dgcr8 -/-) phenotype in ESCs, which is a cell cycle defect leading to an inability to silence self-renewal genes preventing differentation, uncovered a role for let-7 family of miRNAs in silencing self-renewal program upon differentiation [94]. To further integrate miRNAs into the transcriptional circuitry of ES cells Marson et al. used ChIP-Seq of core pluripotency factors combined with quantitative sequencing of short transcripts in ES cells, as well as somatic cells [95]. In agreement with previous studies [76, 78, 79], a pattern of binding and regulation similar to protein coding genes was identified whereby core pluripotency factors occupy the promoters of most miRNAs preferentially or uniquely expressed in ES cells, and miRNA genes are silenced in ES cells by PcG proteins only to be expressed later in development in specific lineages [95].

These studies and others have begun to illustrate the pathways acting downstream of LIF/Stat3 and the genetic hierarchy between different pluripotency factors. Mechanisms ensuring the stability of the transcriptional network of ESCs have afforded them the ability to continue self-renewal in the absence of certain signaling inputs, relying on just a few essential factors for propagation.

### 1.4: Epigenetic regulation of pluripotency and differentiation

The epigenome of ESCs must generate heritable gene expression programs concordant with self-renewal while also retaining the potential to reactivate all developmental programs upon differentiation. This section provides an overview of the various epigenetic hallmarks associated with pluripotency as well as key epigenetic mechanisms that maintain the epigenome of ESCs.

One hallmark of the ES cell epigenome is the X chromosome state that is frequently utilized as a marker for "ground state" pluripotency. In female mouse embryos the paternal X chromosome is silenced (Xi) during early cleavage but the inactive X (Xi) is then reactivated in the ICM, specifically demarcating the nascent epiblast [69, 96, 97]. Shortly thereafter, during uterine implantation, the epiblast transforms into the egg cylinder and in female embryos, one of the X chromosomes undergoes random X inactivation [98]. Female ESCs are characterized by two activated X chromosomes (XaXa) and it is believed this reflects their preimplantation origins, in which female epiblast cells have not yet undergone X inactivation. Murine ESCs that are responsive to LIF/Stat3 and express canonical pluripotency markers such as Oct4, Sox2, Nanog, and Klf4, maintain an X-chromosome state of XaXa and show random X-inactivation following differentiation towards somatic lineages, an epigenetic hallmark that is recapitulated even in ESCs derived from cloned embryos [99]. Recent work has uncovered Nanog as a key marker, and potentially mediator of the domain of X chromosome reactivation within the developing ICM of blastocyst embryos [69]. In addition, Oct4 has been proposed to regulate X chromosome pairing and counting in ESCs [100]. These data reinforce the close association between pluripotency factors and epigenetic regulation of X chromosomes.

Higher order chromatin in ESCs has several features that distinguish it from that of somatic or differentiated cells. Morphologically, ESC chromatin appears as a few dis-

tinct domains of heterochromatin that remain large in size but become smaller, more abundant, and hypercondensed upon differentiation [101-104]. In addition, chromatin-associated proteins, such as the architectural heterochromatin protein HP1 and the linker histone H1, display hyperdynamic or looser binding profiles not found in differentiated cells [105]. Analysis of histone tail modifications associated with euchromatin revealed ESCs have high levels of acetylation at lysine 9 of histone H3 (H3K9ac) as well as methylation of H3K4 at lineage specific genes when compared to embryonic carcinoma cells, hematopoietic stem cells (HSCs), and lymphocytes. These authors suggest lineage-specific genes are primed for expression in ES cells but avoid activation due to opposing chromatin modifications [106].

Indeed, global histone tail modification profiles show unique patterns in ESCs and reveal how epigenetic mechanisms control the undifferentiated state in vitro. In addition to developmental genes necessary for somatic lineages, pluripotency-associated transcription factors also contribute to the repression of genes necessary for non-epiblast lineages of the early embryo, such as the trophoblast or primitive endoderm [24, 55, 62], suggesting global repression of differentiation-inducing factors is a key mechanism in maintaining undifferentiated ESCs in vitro. Many of the repressed targets of Oct4, Sox2, and Nanog show a close association with developmental regulators also bound by Polycomb group (PcG) proteins, which contribute repressive chromatin modifications and gene silencing [107]. PcG proteins comprise at least two distinct repressor complexes (PRC1 and PRC2-PRC3) with core components that are conserved from fly to human [108]. The aforementioned genome-wide studies in human ES cells revealed many of the promoters occupied by PcG proteins also contained the repressive histone modification H3K27me3, often in expansive regions not only occupying gene promoters, but extending from 2 to 35 kb away [78]. The H3K27me3 modification plays a

role in maintaining silencing and undifferentiated state of ESCs as many target genes become de-repressed in the absence of the PCRC2 components Eed or Suz12, leading to precocious differentiation [79, 109]. In fact, PcG-bound genes appear "poised" for activation upon differentiation, as these genes are preferentially activated upon ES cell differentiation [79, 109]. A concept emerging from genome-wide studies of chromatin modifications was specific histone modifications with opposing effects ("silencing" vs. "activating") were frequently found to co-exist at developmental gene loci in ESCs. Large H3K27me3 regions frequently overlap smaller H3K4me3 enriched regions at developmental gene promoters, termed "bivalent domains" [109]. Although largely silent in ES cells, these regions adopt some properties associated with an open chromatin structure, one of which is early replication timing [106]. These observations provide insight into organization of chromatin in pluripotent cells and suggested a model by which silenced developmental genes harboring bivalent promoters remain "poised" for activation upon differentiation.

Whether the bivalent chromatin configuration can be equated with pluripotent chromatin remains to be determined. Recent work using nonmammalian vertebrate embryos has revealed that during the maternal-zygotic transition in Zebrafish more than 80% of genes acquire H3K4me3, with many also marked by H3K27me3, suggesting bivalent chromatin structure is not a tissue culture artifact [110]. However, Xenopus embryos undergoing gastrulation show more spatially regulated expression of genes that correlates with subsequent deposition of H3K27me3, whereas H3K4me3 correlates with earlier, zygotic gene activation [111]. Therefore the in vivo function of opposing histone modifications remains under investigation. It is worth noting several alternative models have been proposed as well. First, based on findings that H3K4me3 is found at most protein coding genes in human embryonic stem cells with a stalled RNA PolII, Young

and colleagues proposed that regulation of many developmental genes might be at the level of elongation [112]. Moreover, there is increasing evidence dynamic gene expression changes, as well as alternations to the epigenome, occur upon isolation of ESCs that may not necessarily reflect the epiblast in vivo. Recent work using single-cell RNA profiling of ICM-outgrowths has revealed over half of the known epigenetic modifiers change their expression during ICM outgrowth with repressive regulators being most highly expressed [113, 114]. The extent to which modifications of chromatin in ESCs reflects an in vitro adaptation also continues to be an area of intense investigation.

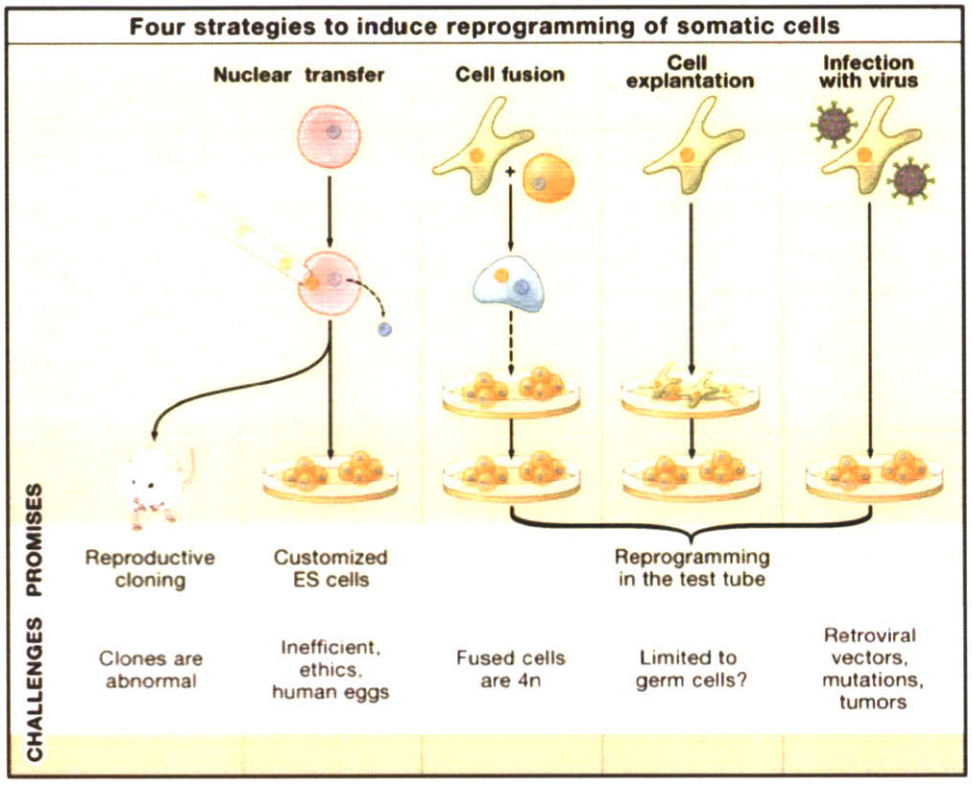
DNA methylation is an essential epigenetic modification required for development [115-118]. Although it is unnecessary for the propagation of ESCs in vitro these cells cannot differentiate[118]. Reduced representation bisulphite sequencing [119, 120] has been applied to both embryonic and somatic cells to analyze genome-wide patterns of DNA methylation in the context of CpG dinucleotides. CpG islands, found at about 60% of human gene promoters, revealed a unique pattern in ESCs depending on the density of CpGs within the particular regions [121]. High-CpG-density promoters (HCPs) are mostly unmethylated and are associated with 'housekeeping' genes and highly regulated developmental genes. Virtually all HCPs are marked by H3K4me3 in ESCs, regulated by trithorax-group (trxG) proteins, while others exhibit regulation by trxG and PcG proteins reflected in a bivalent chromatin state (H3K4me3 and H3K27me3). HCPs marked only by H3K4me3 tend to retain this mark even following differentiation, whereas bivalent chromatin states tend to lose one or both marks with a small fraction remaining bivalent [121, 122]. Hypermethylation of CpG-dense regions lead to exclusion of trxG/PcG activity and long-term irreversible gene silencing. A second class, termed low-CpG-density promoters (LCPs), is largely methylated with only a subset (between 3-7%) containing euchromatin-associated histone modifications such as H3K4me1/2 and reduced

methylation levels in ESCs. In contrast to CpG-rich regions, periods of methylation and de-methylation can occur during differentiation, which may reflect the subset of cis-regulatory elements that exist within this pool of regulated DNA, among other non-coding elements such as transposons or interspersed repeat elements [121]. This data suggests DNA methylation at distinct regions (CpG-rich and –poor) is regulated differently in ESCs and during differentiation. Interestingly, analysis of ESC-derived multipotent neural progenitors and primary cells demonstrated aberrant hypermethylation during extended proliferation in vitro at some 'weak' CpG islands at particular developmentally regulated genes, again reinforcing that caution should be taken in correlating observations with cells grown in vitro to their in vivo counterparts [121, 123].

**In vitro reprogramming of somatic cells to pluripotency**

As described in the previous sections the molecular and epigenetic factors influencing pluripotency are beginning to be understood. Not only do ESCs hold great promise for understanding cellular differentiation but also for medical purposes such as patient-specific regenerative medicine. The isolation and derivation of embryonic stem cell lines however, either from fertilized or cloned embryos, requires destruction of the embryo raising ethical as well as practical questions that have hindered progress towards this goal. Alternative sources of pluripotent cells such as reprogramming from adult somatic cells, avoids such constraints and has remained a major focus for the stem cell community. This section will describe strategies that have been employed to generate pluripotent cells in vitro while emphasizing their impact in addressing questions surrounding reprogramming mechanisms generally (**Figure 5**). However, to date many of these approaches still have limitations that preclude use in broader applications of regenerative medicine.

**Figure 5: Four strategies to induce reprogramming of somatic cells** (Jaenisch and Young, 2008)




One such strategy tested whether epigenetic reprogramming of somatic nuclei can be achieved following fusion of somatic cells with pluripotent embryonic cells. The general conclusion to date has been that fusion products exhibit the phenotype of the less-differentiated fusion partner. Fusion of pluripotent teratocarcinoma cells with primary thymocytes resulted in the formation of pluripotent hybrids that gave rise to tumors [124]. Following similar protocols using murine, and later human ESCs, cell hybrids exhibit many characteristics of ESCs confirming that reprogramming factors exist within these cells in addition to the egg and early cleavage blastomeres. Functional as well as molecular tests were performed in order to determine the extent of reprogramming of the somatic cell chromosomes, as one formal possibility is they are simply pas-

sengers in hybrid cells. However several observations support the conclusion that extensive reprogramming of somatic chromosomes occurs following cell fusion. At the molecular level hybrids show demethylation of CpG-containing promoters of pluripotency genes with re-activation of Oct4, as well as reactivation of the inactive somatic X chromosome in female lymphocyte-ES-cell hybrids [125, 126]. At the functional level, teratomas could be generated following subcutaneous injection into SCID mice, and genes of all three germ layers were expressed [127]. From a mechanistic viewpoint reactivation of pluripotency genes has been proposed to require cell division with the first detectable marker appearing after 2 days [128], while Nanog transgenic ESCs induce pluripotency in somatic cell hybrids at a much greater efficiency than non-transgenic ESCs upon fusion with neural stem cells[129], suggesting core pluripotency factors can enhance the reprogramming process. Yet there is limited evidence that fusion induces extensive genome-wide reprogramming in somatic cells and careful genome-wide analysis of epigenetic marks on somatic chromosomes has not been performed to date. In addition the generation of tetraploid cells without the ability to eliminate chromosomes generating diploid pluripotent cells, precludes the application of this technology in regenerative medicine.

A second method to isolate pluripotent cells in vitro is by culture-induced reprogramming and seems to be confined to a few donor cell types originating from the germ lineage. Primordial germ cells (PGCs) emerge at gastrulation and, in male embryos, give rise to spermatogonial stem cells. Both of these cell types have been shown to be capable of generating pluripotent cells by explanting the cells from embryos at different stages of development. Growth conditions are essential for the reprogramming of these cells. From embryonic day 8.5 embryos, primordial germ cells grown in vitro give rise to embryonic germ (EG) cells [130]. EG cells require LIF/Stat3 for growth, express all canonical pluripotency markers, give rise to teratomas with cells of all three germ layers,

and can generate adult mouse chimeras [130]. Spermatogonial stem cells from newborn and adult male gonads also generate ES-like cells, although at very low efficiency and after an extended time in culture. Termed male germ stem cells (maGSCs), these ESC-like cells can be propagated in serum and LIF and uniquely carry a male-specific imprinting pattern, also capable of inducing teratomas and contributing to chimeras with germline transmission [131, 132]. While these studies reveal a source of potential adult somatic cells capable of generating pluripotent cells in vitro, evidence that similar cell types exist in human capable of undergoing reprogramming as described for murine cells is still lacking.

**Direct reprogramming to pluripotency with defined factors**

The aforementioned techniques applied towards the isolation of pluripotent cells from adult somatic cell donors all suffer from some form of technical, ethical, or practical constraints that preclude their broad use in generating patient-specific pluripotent stem cells. Yet another alternative was the use of individual transcription factors to perform direct conversion from one cell type into another. Weintraub and colleagues attempted "transdifferentiation," or conversion of one somatic cell to that of another somatic cell without transiting through an embryonic intermediate, turned fibroblasts into muscle-like cells by over-expression of the master regulator of muscle differentiation, MyoD [133, 134]. While some muscle-specific gene activation occurred, re-activation of the endogenous *MyoD* gene was not detected in the converted muscle cells, suggesting that epigenetic reprogramming was incomplete and that maintenance of the myogenic phenotype depended on the transgene. Indeed, the *MyoD* promoter continued to be methylated even after the conversion, suggesting incomplete epigenetic reprogramming. These studies however, emphasized the power of individual transcription factors to ac-

tivate target genes epigenetically silenced in cells from which they are not normally expressed.

In 2006 Takahashi and Yamanaka altered the protocol described above only slightly but achieved dramatically different results by reprogramming somatic cells back to a pluripotent ES-like state with four transcription factors [135]. Starting with a pool of 24 candidates, some of which were regulators of pluripotency (e.g. Nanog, Oct4, Sox2, Stat3, Rex1, Klf4), others that were factors thought to facilitate opening of chromatin (e.g. c-Myc), a screen was performed by retroviral transduction using genetically modified embryonic or adult murine fibroblasts carrying a drug-selectable reporter at the endogenous promoter of an Oct4 target, *Fbx15*. Neomycin was added to culture medium between 3-9 days after transduction of the factors to select for cells that had reactivated the *Fbx15* locus. Indeed colonies appeared within about 9-12 days and could be mechanically passaged at around two weeks and grown into stable cell lines thereafter. With initial success using a pool of all 24 factors, single factors were eliminated one-by-one, ultimately revealing that four factors (Oct4, Sox2, Klf4, and c-Myc) were sufficient to induce ESC-like cells from adult somatic cells, coined induced pluripotent stem cells (iPSCs) [135]. Cells that had reactivated *Fbx15* were shown to be pluripotent by their ability to form teratomas and contribute to somatic tissues in embryonic chimeras, however live chimeras were not obtained. These cells had reactivated many pluripotency genes but importantly the endogenous *Oct4* and *Nanog* genes were not expressed or were expressed at a lower level than in ES cells, with their respective promoters largely methylated. Therefore although these cells had many properties of ES cells, they most likely reflected an "incompletely" reprogrammed state.

Building on these studies Yamanaka and two other groups utilized more stringent criteria to isolate fully pluripotent iPS cells by selecting for reactivation of the en-

dogenous *Nanog* or *Oct4* genes [136-138]. Although extremely inefficient, ~0.001-0.1% of infected cells give rise to iPSCs, by genetic, epigenetic, and developmental criteria the cells reflected an authentic pluripotent state: (i) global gene expression of iPS cells showed highly similar profiles to ES cells; (ii) the chromatin configuration of *Oct4* or *Nanog* genes was indistinguishable from ES cells (H3K4me3) and the endogenous Oct4 and Nanog promoters were hypomethylated; (iii) X-chromosome status had been reprogrammed to an ESC-like state of XaXa in female iPS cells; (iv) many developmental and lineage-specific genes had acquired 'bivalent' epigenetic chromatin marks at their promoters; (v) iPS cells generated post-natal chimeras and contributed to the germ line with some lines capable of giving rise to late gestation embryos through tetraploid complementation [136], the most stringent test for developmental potency.

An important question was whether genetic modification of the donor cells was necessary to identify and isolate fully reprogramming iPS cells, which would prove difficult in human somatic cells as the protocol involved drug-selectable or green flourescent protein (GFP) reporters knocked-in to the endogenous loci of *Oct4* or *Nanog*. The successful isolation of genetically unmodified murine iPSCs by morphology alone argued these concerns could be overcome [139]. Another outstanding question pertained to the inability of most iPS cell lines to give rise to liveborn adult "all iPSC" mice in tetraploid complementation assays, with most studies describing only derivation of late gestation "all iPSC" embryos (~E12-14.5). This protocol is the most stringent test of pluripotency and forces iPSCs to give rise to all tissues of the adult animal. Although still not completely settled, a few recent reports have confirmed the derivation of iPSCs capable of generating liveborn "all iPSC" pups and adult mice [140-142], arguing that reprogramming can, in some cases, give rise to fully pluripotent stem cells with all properties

of ICM-derived embryonic stem cells. Whether technical or biological constraints are behind the limited isolation of 4n-competent iPSCs remains under investigation.

Reprogramming by defined factors is a slow and inefficient process compared to SCNT or cell fusion, where reactivation of Oct4 or Nanog occurs within 1-2 cell divisions. During in vitro reprogramming to generate iPSCs reactivation of core pluripotency genes appears to take weeks. Retroviral delivery of reprogramming factors allowed endogenous pluripotency genes to take over control of the pluripotent gene expression program as retroviruses become silenced by DNA methylation in embryonic cells, as was observed in Nanog- or Oct4-selected iPSCs [136-138]. Drug-inducible viruses of Oct4, Sox2, Klf4, and c-Myc, showed a minimum time of 9 days ectopic expression is required for generation of stable iPSCs in mouse [143, 144].

Shortly after the derivation of germline-competent murine iPSCs, success with human somatic cells gave rise to human induced pluripotent stem cells (hiPSCs) [145-147]. One group used the previously described set of transcription factors but alternatives to the reprogramming factor cocktail being were reported as well (Oct4, Sox2, Nanog, and Lin28) [147]. Yet the demonstration that defined factors could reprogram cell types other than fibroblasts as well as terminally differentiated cells remained unknown. Yamanaka and colleagues reported the generation of iPSCs from adult mouse liver and gastric epithelial cells [148] and the successful reprogramming of terminally differentiated B-cells provided definitive evidence that this protocol can be applied to a wide range of cell types other than fibroblasts [149]. Importantly, the ability to induce pluripotency with the combination of Oct4, Sox2, Klf4, and c-Myc appears to be evolutionarily conserved among mammals as iPSCs from somatic cells of species other than mouse and human have now been derived including rat, pig, sheep, rabbit, bovine, and monkey.

Previously SCNT and gene therapy had been had been combined to repair diseased lineages in mice using cloned ESCs to derive genetically repaired hematopoetic cells [150]. In a similar effort, iPSCs proved capable of repairing diseased tissues in several studies using mouse models of sickle cell anemia and Parkinson's [151, 152]. Moreover with the derivation of hiPSCs efforts turned towards the generation of disease-specific human iPSCs [153-155] with the ability to employ differentiation protocols to examine the molecular profile of the diseased cell type in vitro.

Derivation of pluripotent cells using direct reprogramming with defined factors represented a breakthrough towards generating personalized pluripotent stem cells without the ethical or practical constraints associated with embryonic stem cells. However technical limitations remained in that the reprogramming process involved integration of proviruses carrying oncogenes such as Klf4 and c-Myc, with some iPSCs harboring between 15-20 total proviral integrations. Indeed tumors have been reported from chimeric mice generated from iPSCs due to reactivation of the c-Myc proviruses [156]. Extensive efforts to eliminate the use of viruses or integrating DNA elements have been underway for several years with successes being reported in both mouse and human reprogramming.

First, the recipe(s) for reprogramming revealed several factors were not genes exclusive to ESCs, and suggested that potentially somatic cell types that expressed Sox2, Klf4 or c-Myc, could be used to eliminate the need for additional delivery of transgenes carrying these factors. To this end, multipotent neural progenitor cells (NSCs) proved successful in both mouse and human. Elimination of all factors except Oct4 has been reported, generating either 2-factor (Oct4+Klf4 or Oct4+c-Myc in mouse), or 1-factor iPSCs (Oct4 only in both mouse and human) [157-159]. Interestingly however, similar experiments reported by others suggested in NSCs reprogramming efficiencies were

lower when Sox2 was delivered ectopically in addition to three other factors suggesting the relative protein stochiometry of reprogramming factors could influence the generation of iPSCs [160]. Although successful in eliminating some factors, use of somatic cells expressing endogenous levels of reprogramming factors has not yet overcome the need for ectopic delivery of Oct4.

Second, success has been reported in generating integration-free iPSCs mostly in human and mouse fibroblasts with many different delivery mechanisms: 1) *repeated infection/transfection:* adenoviruses carrying Oct4, Sox2, Klf4, and cMyc or by repeated transfection of two plasmids carrying the same factors, or in human, repeated transfection of multiple episomal vectors carrying seven reprogramming factors [161-163]; 2) *excisable viruses*: including a single polycistronic vector carrying reprogramming factors leading to one proviral integration that is excised upon isolation of iPSCs [154, 164, 165]; 3) *protein delivery:* purified Oct4, Sox2, Klf4 and cMyc [166]; and 4) *modified RNA delivery:* Oct4, Sox2, Klf4, and cMyc, to human fibroblasts to generate hiPSCs [167]. Although technically iPSCs can be generated using all of the aforementioned protocols, with the recently reported RNA-delivery being most promising, most remain highly inefficient and require weeks if not months to isolate genetically unmodified pluripotent cells.

Third, chemically defined protocols for reprogramming have successfully replaced some factors individually but no such single, or combination of compounds is capable of generating iPSCs without requiring transgenes of at least two reprogramming factors [168-171]. Importantly, the action of some of these compounds has been reported to act by inducing Nanog expression, or by inhibiting chromatin-associated proteins that maintain repressive epigenetic marks at the promoter of Oct4 [168, 171].

**Molecular mechanisms governing in vitro reprogramming by defined factors**

Direct reprogramming to pluripotency has opened up a major biomedical research field with goals of generating patient-specific hiPSCs for regenerative medicine. Yet this technology also generated an alternative approach in studying the underlying mechanisms of epigenetic reprogramming that could not be analyzed by the labor-intensive methods such as SCNT. This section will focus on what is known about the mechanisms of reprogramming by defined factors and will also describe novel tools that have been generated to examine the biological principles of epigenetic reprogramming to pluripotency. Five general observations described herein can be summarized as follows: 1) the action of individual reprogramming factors (Oct4, Sox2, Klf4, c-Myc) during iPSC generation can be replaced by related or unrelated transcription factors, with the former being well documented; 2) when controlling technical parameters associated with in vitro growth and entry into the cell cycle, virtually all somatic cells have the *potential* to give rise to iPS cells given enough cell divisions; 3) reprogramming follows a stochastic course to pluripotency with some cells reprogramming earlier than others but the outcome of any given cell cannot be identified *a priori*; 4) reprogramming is amenable to acceleration with some factors influencing iPSC formation in a cell-cycle-dependent or cell-cycle-independent fashion; 5) genetic or environmental perturbations that negatively or positively influence the stability of ESCs, also affect reprogramming efficiencies and the ability to isolate stable iPSCs.

The four reprogramming factors described by Takahashi and Yamanaka (Oct4, Sox2, Klf4 and cMyc) are members of a family of homologous transcription factors. Oct4 for example, belongs to the octamer-binding (Oct) family of transcription factors, which contain the POU domain [172], and has two closely related homologs Oct1 and Oct6. Similarly, Sox2, part of the SRY-related HMG-box (Sox) transcription factors, characterized

by the presence of the high-mobility group (HMG) domain has many homologs (Sox1, Sox3, Sox7, Sox15, Sox17, Sox18). To test the stringency of reprogramming and whether closely related family members could substitute for reprogramming factors, Nakagawa et al. (2008) performed reprogramming with 13 related transcription factors and found no other substitute for Oct4, however related factors to Sox2, Klf4, and cMyc were capable of generating iPSCs [156]. In addition these authors and others uncovered the dispensability of c-Myc during reprogramming, which seems only to increase efficiencies but not be strictly required [156, 173]. Mice generated from iPSCs show tumors that reflect reactivation of the previously silenced c-Myc provirus, and elimination of this factor from the reprogramming cocktails reduced the rate of tumor formation in some studies [156, 174]. Alternate combinations of known pluripotency factors have now been applied to iPSC generation such as Esrrb, Nr5a2 [175, 176], which are capable of replacing Klf4 and Oct4 in four-factor reprogramming cocktails respectively. The conserved DNA-binding properties of the pluripotency factors and their related family members capable of reprogramming to pluripotency, most likely arises from the ability bind consensus motifs that reactivate the autoregulatory and feed-forward loops by activating endogenous pluripotency genes [76].

When considering reprogramming efficiencies an important issue that still remains today is whether different groups use the same "marker" to indicate a successfully reprogrammed cell. Some markers are less stringent than others. Alkaline phosphatase (AP) and SSEA1 are not as exclusive to ESCs as are Nanog or Oct4. The use of reactivation of endogenous *Oct4* or *Nanog* by GFP knock-in reporters appears to most faithfully reflect a fully reprogrammed state. Using reactivation of Oct4 or Nanog as detected by GFP, reprogramming efficiency in various studies on mouse cells have indi-

cated that approximately 0.001-0.5% of cells infected with all four factors are reprogrammed into iPS cells after approximately 2-3 weeks [136, 138, 139].

A system of doxycycline-inducible lentiviruses was created that allowed temporal control of reprogramming factor expression, which also had advantages to retroviruses that become irreversibly silenced upon entry into the pluripotent state [143, 144]. During reprogramming, non-specific, pluripotency-associated markers such as alkaline phosphatase (AP) and SSEA1 appear in a significant fraction of cells only a few days after induction of reprogramming factors while a small fraction express the core pluripotency factors such as Sox2, Nanog, or Oct4 late in the initial formation of iPSCs (~ d14-16). Experiments utilizing the temporal control of virus expression, by withdrawl of doxycycline (DOX) from the culture media at various times during reprogramming, demonstrated the minimum time required to generate mouse iPSCs with Oct4, Sox2, Klf4, and c-Myc is about 9 days and efficiencies increasing with prolonged exposure to reprogramming factors [143, 144]. Other markers of the pluripotent state were examined such as the ability to silence retroviruses as well as X-reactivation. Using somatic cells with a Sox2 GFP knock-in reporter to observe reactivation of endogenous loci combined with infection with a retroviral Tomato reporter, the pattern of expression between fluorescent reporter and GFP was mutually exclusive, suggesting the ability to silence closely correlated with the reactivation of core pluripotency factors [144]. Similar observations were made using X-linked GFP reporters suggesting reactivation of X-chromosomes occurs simultaneously with reactivation of pluripotency markers such as Nanog and Sox2 [144].

A hypothesis for such low reprogramming efficiencies was that upon initial transduction with reprogramming factors only a small fraction of infected cells received the appropriate proviral integration patterns that led to a specific expression level and

protein stochiometry that was optimal for iPSC generation. This was potentially reflected in the variation in the kinetics of reactivation of pluripotency markers such as *Nanog* or *Oct4* during reprogramming upon primary infections and could not be ruled out because of the genetically heterogenous population of cells undergoing reprogramming. This lead to a prediction that if somatic cells had a similar integration pattern and presumably similar expression of reprogramming factors they should reprogram with similar efficiencies and kinetics.

A key insight made by Meissner et al. (2007) was that even when controlling for integration patterns, reprogramming depends on stochastic epigenetic events. Clonal analyses demonstrated that the activation of pluripotency markers can occur at different times in individual mitotic daughter cells of the same infected fibroblast [84, 139], suggesting cell divisions give rise to subclones with different biological properties (**Figure 4**). This most likely reflects epigenetic changes in chromatin modifications or DNA methylation that occur differently between clones.

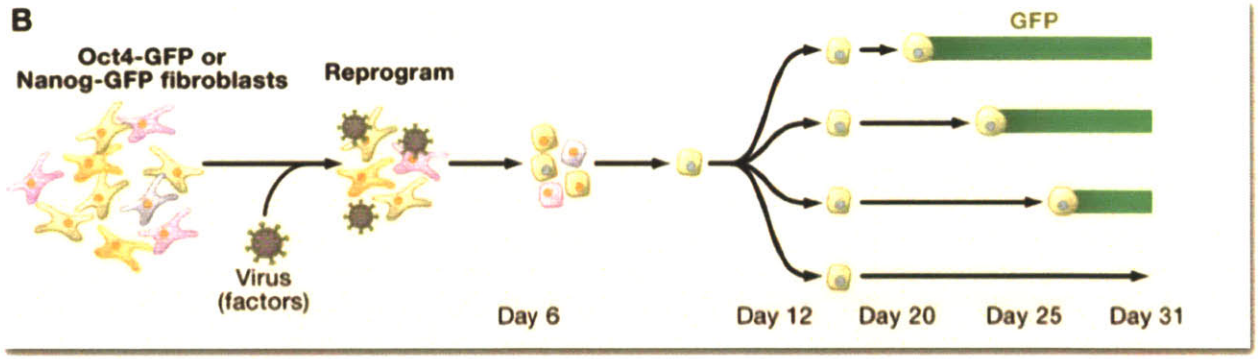



**Figure 4: Direct reprogramming involves stochastic epigenetic events**

Indeed both DNA methylation and histone modifications have been reported to influence reprogramming efficiencies. Use of the DNA methyltransferase inhibitor 5-azacytidine during reprogramming increased reprogramming efficiencies most likely by inhibiting the maintenance of repressive DNA methylation at key pluripotency

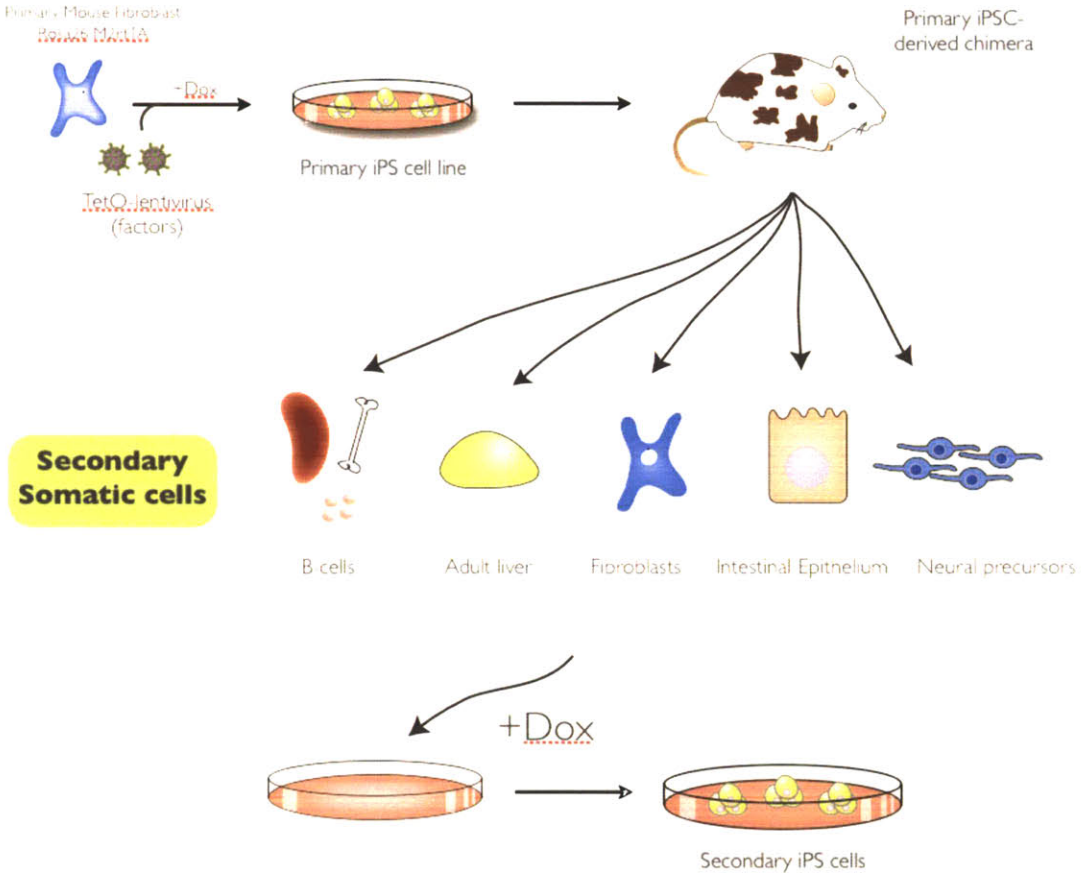
genes [177]. In addition, histone deacetylase (HDAC) inhibitors such as valproic acid (VPA) improved reprogramming efficiencies and enabled derivation of 2-factor human iPSCs only requiring infection with Oct4 and Sox2 [178, 179]. Klf4 and c-Myc may solely be responsible for "priming" chromatin during reprogramming, enabling more direct access to Oct4 and Sox2 targets as pre-infection with the former reprogramming factors followed by infection with the latter, increased the kinetics of iPSC formation [180].

Given that growth and propagation of ESCs is positively or negatively regulated by cytokines or small molecules added to the medium, increased reprogramming efficiencies have been reported by enhancing Wnt signaling or use of "2i/LIF" conditions [181, 182], suggesting similar pathways regulate the establishment of iPSCs. Importantly, inhibitors of pluripotency in ESCs are similarly observed to affect iPSCs generation. In agreement with previous studies using Nanog null ESCs, iPSCs cannot be established from somatic cells in which Nanog is disrupted underscoring the essential role Nanog plays in *entry* into the pluripotent state [69].

The drug-inducible lentivirus systems gave way to the "reprogrammable mouse model" for studying induction of pluripotency in homogeneous transgenic somatic cells (**Figure 6**). In this "secondary system" strategy, primary iPS cells are generated from fibroblasts infected with Dox-inducible lentiviral vectors carrying the reprogramming factors Oct4, Sox2, Klf4, and c-Myc. Single iPS clones are injected into mouse blastocysts to generate chimeras and contribute to a variety of genetically homogeneous "secondary" somatic cells that carry the same Dox-inducible factors in their genome as their primary iPS cells. Cells from chimeric mice created using genetically reprogrammed cells can be triggered via drug administration to enter an ES-cell-like state without the need for additional direct genetic manipulation. This technical advancement enables creation of large numbers of genetically identical somatic cells that can be repro-

grammed to an ES-cell-like state simply by exposure to a drug. Because the injected iPSC lines carried a constitutively expressed antibiotic resistance gene, homogenous iPSC-derived somatic cell populations such as embryonic fibroblasts, mesenchymal stem cells, neural precursors, and lymphocytes could be isolated that carried provirus integration patterns identical to those in the primary iPS cell lines [183, 184].

**FIGURE 6: Reprogramable mouse strain**




Predictions of increased efficiencies once genetically identical somatic cells were tested did bear out as secondary systems could now generate iPSCs between 1-10% after 2-3 weeks [183, 184]. These data argued that appropriate factor expression levels influence reprogramming however major epigenetic barriers must be overcome because the majority of cells still do not generate iPSCs after 2-3 weeks. In addition not all secon-

dary systems proved equivalent, each had unique induction levels of the reprogramming factors in somatic tissues of reprogrammable mice. In some lines certain proviruses could not be reactivated in all somatic tissues tested because they had becoming epigentically silenced during differentiation towards particular somatic lineages. For example, in one reported secondary line Sox2 was readily activated in several somatic tissues such as neural precursors or keratinocytes but was unable to be reactivated in intestinal epithelium cells precluding the generation of iPSCs until Sox2 was reintroduced [184].

A second advantage of the secondary systems was that germline transmission of iPSC genomes carrying dox-inducible proviruses allowed screening for segregation of some reprogramming factors to generate mice carrying different combinations of proviruses to use as a drug-screening platform for the missing factors [180]. However because this system was based on infection with individual viruses, up to 7 proviral integrations had to be tracked with each generation therefore it proved extremely difficult and laborious to maintain a mouse colony containing all four factors.

A landmark study using secondary systems was performed by long-term analyses of clonal cell populations derived from pro/pre-B cells of the hematopoetic lineage to determine the dynamics of reprogramming and to assess the fraction of donor cells that are susceptible to reprogramming [185]. These populations have a higher cloning efficiency for single cells and better tolerance of reprogramming factors. In addition, these cells represent a well-defined lineage-committed population with the rearrangement of the IgH locus allowing for the unambiguous retrospective identification of the donor cells. In this study reprogramming efficiency was defined as the *potential* of a donor cell to generate an iPSC daughter at some point. Plating single secondary pro/pro-B cells into single wells and subsequently passaging until the apprearance of iPSCs as indi-

cated by a Nanog-GFP reporter, nearly *every* cell was able to generate iPSCs (>90%) [185]. The kinetics of reprogramming displayed a broad distribution for isolation of iPSCs, spanning 2-18 weeks, consistent with the observation that the process involves stochastic and rate-limiting epigenetic event(s). Additional inhibition of the p53/p21 pathway or ectopic expression of Lin28, while controlling for the same growth conditions, accelerated conversion into iPSCs and was directly proportional to the increase in cell division rate. Thus, p53 inhibition did not increase the fraction of cells that could be reprogrammed but rather accelerated the formation of iPSCs in time, which appeared after a similar number of cell divisions. In contrast to these findings, ectopic expression of Nanog accelerated reprogramming in a manner that was independent of the rate of cell divsions [185, 186].

**Does the induced pluripotent state represent ESC-like pluripotency?**

An unresolved question in the reprogramming field is whether iPSCs are equivalent to ESCs. If there are differences such as extensive genetic or epigenetic abnormalities capable of influencing iPSCs during differentiation and/or transplantation this could preclude the application of this promising technology in regenerative medicine. Yamanka's group first reported that iPSCs derived from different adult tissues showed variation in function compared to ESCs both in vitro and in vivo. First, some iPS cell lines were unable to form primary neurospheres in standard in vitro assays, and after generating secondary neurospheres (SNS) from additional iPS cell lines, these SNSs showed different behaviors *in vivo*, either incorporating into brain tissue or developing large teratomas after injection into NOD/SCID mice [174]. Of the 31 mice injected with ESC-derived SNSs (28/31 or 90%) lived for up to 45 weeks with only one revealing a small tumor after autopsy. In contrast a significant fraction of the mice injected with iPSC-derived SNS either died prematurely with some having observable tumors, while

others that survived through the experimental period, had tumors of varying sizes upon dissection. Strikingly, some lines exhibited mortality rates higher than 80%. These issues raised valid concerns about the safety of differentiated tissues from iPSCs. Additional reports suggest the cellular physiology of iPSC-derived somatic cells is limited and that reprogramming causes potentially harmful genomic alternations that may consequently lead to tumors or other cellular defects upon differentiation and transplantation in vivo. Careful analysis of mouse and human iPSCs and ESCs has identified recurrent chromosomal aberrations including a high incidence of chromosome 8 (mouse), or 12 and 17 duplications (human), while some hiPSC-derived somatic cells such as hemangioblastic derivatives, undergo early senescence and show limited expansion in vitro [187].

Compared to functional tests for pluripotency, molecular analyses allow for more quantitative comparisons of iPSCs and ESCs. Some initial investigation into iPSCs derived from multiple adult tissues suggested that at least for some clones, iPSCs are similar if not indistinguishable from ESCs derived from embryo of nuclear transfer experiments. These included profiling of global gene expression, modifications of histone tails, the state of X chromosome inactivation, and profiles of DNA methylation. However, some studies using global expression analyses concluded that iPSCs are a unique subtype of pluripotent cells that retain a consistent gene expression signature distinguishable from ESCs, even after extended passaging [188]. These results have been reanalyzed and new datasets generated by other groups who have reported that most likely small variations between human iPSCs and ESCs in chromatin structure and global gene expression may constitute experimental "noise" and do not reflect a consistent signature that distinguishes iPSCs from ESCs[189] [190].

Assays for developmental potency in mice, such as chimera formation and germline contribution are routinely passed in many of the iPSCs derived to date. However the production of "all-iPSC mice" by tetraploid complementation (4n) has proven more difficult for the vast majority of iPSCs. However as stated previously some success in generating tetraploid-competent iPSCs has been reported. Two recent studies concluded that the only distinguishable difference between ESCs and the vast majority of iPSCs was the abnormal reduction in the expression of the maternally imprinted Dkl1-Dio3 locus and that this expression difference was the underlying cause for the inability to generate all-iPSC mice by tetraploid complementation. Another study compared the patterns of global DNA methylation and in vitro differentiation of early passage iPSCs derived from B lymphocytes or fibroblasts with those of ESCs from fertilized embryos or by nuclear transfer. Analysis of the methylation profiles suggested that reprogramming with transcription factors frequently creates iPSC lines that have retained epigenetic marks reminiscent of their cell-of-origin. Others report similar results but observe these phenotypes reside with extended passaging [191]. These data argue differences in gene expression, chromatin marks and DNA methylation endow iPSC lines with, depending on your perspective, functional limitations represented by a unique ability to give rise to somatic cells of their tissue-of-origin more efficiently than other lineages, a phenomenon termed epigenetic memory [192]. However some lines showed only partial demethylation of the endogenous promoter of Nanog, a key "licensing" factor required for entry into pluripotency, therefore it could be argued these cells were incompletely reprogrammed and comparisons to ESCs or other iPSCs that are completely reprogrammed would not yield reliable results.

Many parameters unfortunately have not been controlled for in these studies. These include the presence and incomplete silencing of transgenes, different combina-

tions of reprogramming factors used to induce iPSCs, natural heterogeneity that exists between different pluripotent ESC lines, incomplete reprogramming in early passage cell lines, and the genetic background of cells. In addition many of the reported phenotypes of iPSCs can also be found in ESCs, such as subtle but detectable levels of genomic abberations [193]. Therefore the key question is whether *any* of the described phenotypes lie outside the variation that exists between ESCs. Careful control for many parameters described above will allow for more definitive tests of the similarities between ESCs and iPSCs in developmental assays.

REFERENCES CITED


1. Batlle-Morera, L., Smith, A. & Nichols, J. Parameters influencing derivation of embryonic stem cells from murine embryos. *Genesis* **46**, 758-767 (2008).

2. Gardner, R.L. & Brook, F.A. Reflections on the biology of embryonic stem (ES) cells. *Int J Dev Biol* **41**, 235-243 (1997).

3. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156 (1981).

4. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-7638 (1981).

5. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147 (1998).

6. Briggs, R. & King, T.J. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc Natl Acad Sci U S A* **38**, 455-463 (1952).

7. Gurdon, J.B. & Byrne, J.A. The first half-century of nuclear transplantation. *Proc Natl Acad Sci U S A* **100**, 8048-8052 (2003).

8. McGrath, J. & Solter, D. Nuclear transplantation in the mouse embryo by microsurgery and cell fusion. *Science* **220**, 1300-1302 (1983).

9. McGrath, J. & Solter, D. Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro. *Science* **226**, 1317-1319 (1984).

10. Wells, D.N., Misica, P.M., Day, T.A. & Tervit, H.R. Production of cloned lambs from an established embryonic cell line: a comparison between in vivo- and in vitro-matured cytoplasts. *Biol Reprod* **57**, 385-393 (1997).

11. Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J. & Campbell, K.H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810-813 (1997).

12. Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369-374 (1998).

13. Egli, D., Rosains, J., Birkhoff, G. & Eggan, K. Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447**, 679-685 (2007).

14. Egli, D., Sandler, V.M., Shinohara, M.L., Cantor, H. & Eggan, K. Reprogramming after chromosome transfer into mouse blastomeres. *Curr Biol* **19**, 1403-1409 (2009).

15. Cibelli, J.B. et al. Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science* **280**, 1256-1258 (1998).

16. Onishi, A. et al. Pig cloning by microinjection of fetal fibroblast nuclei. *Science* **289**, 1188-1190 (2000).

17. Polejaeva, I.A. et al. Cloned pigs produced by nuclear transfer from adult somatic cells. *Nature* **407**, 86-90 (2000).

18. Baguisi, A. et al. Production of goats by somatic cell nuclear transfer. *Nat Biotechnol* **17**, 456-461 (1999).

19. Wakayama, T. & Yanagimachi, R. Cloning of male mice from adult tail-tip cells. *Nat Genet* **22**, 127-128 (1999).

20. Wakayama, T., Rodriguez, I., Perry, A.C., Yanagimachi, R. & Mombaerts, P. Mice cloned from embryonic stem cells. *Proc Natl Acad Sci U S A* **96**, 14984-14989 (1999).

21. Rideout, W.M., 3rd et al. Generation of mice from wild-type and targeted ES cells by nuclear cloning. *Nat Genet* **24**, 109-110 (2000).

22. Eggan, K. et al. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc Natl Acad Sci U S A* **98**, 6209-6214 (2001).

23. Tsunoda, Y. & Kato, Y. Full-term development after transfer of nuclei from 4-cell and compacted morula stage embryos to enucleated oocytes in the mouse. *J Exp Zool* **278**, 250-254 (1997).

24. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391 (1998).

25. Bortvin, A. et al. Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development* **130**, 1673-1680 (2003).

26. Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035-1038 (2002).

27. Eggan, K. et al. Mice cloned from olfactory sensory neurons. *Nature* **428**, 44-49 (2004).

28. Inoue, K. et al. Generation of cloned mice by direct nuclear transfer from natural killer T cells. *Curr Biol* **15**, 1114-1118 (2005).

29. Jaenisch, R. Human cloning - the science and ethics of nuclear transplantation. *N Engl J Med* **351**, 2787-2791 (2004).

30. Bradley, A., Evans, M., Kaufman, M.H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-256 (1984).

31. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**, 2048-2060 (1998).

32. Ernst, M., Oates, A. & Dunn, A.R. Gp130-mediated signal transduction in embryonic stem cells involves activation of Jak and Ras/mitogen-activated protein kinase pathways. *J Biol Chem* **271**, 30136-30143 (1996).

33. Matsuda, T. et al. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J* **18**, 4261-4269 (1999).

34. Stewart, C.L. et al. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76-79 (1992).

35. Ying, Q.L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-292 (2003).

36. Kunath, T. et al. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902 (2007).

37. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**, 183-186 (2003).

38. Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev Biol* **210**, 30-43 (1999).

39. Chen, S. et al. Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci U S A* **103**, 17266-17271 (2006).

40. Ying, Q.L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523 (2008).

41. Cai, L. et al. Promoting human embryonic stem cell renewal or differentiation by modulating Wnt signal and culture conditions. *Cell Res* **17**, 62-72 (2007).

42. Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D. & Niwa, H. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. *Biochem Biophys Res Commun* **343**, 159-166 (2006).

43. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A.H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**, 55-63 (2004).

44. Singla, D.K., Schneider, D.J., LeWinter, M.M. & Sobel, B.E. wnt3a but not wnt11 supports self-renewal of embryonic stem cells. *Biochem Biophys Res Commun* **345**, 789-795 (2006).

45. Nichols, J., Silva, J., Roode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215-3222 (2009).

46. Buehr, M. et al. Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**, 1287-1298 (2008).

47. Li, P. et al. Germline competent embryonic stem cells derived from rat blastocysts. *Cell* **135**, 1299-1310 (2008).

48. Nichols, J. et al. Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. *Nat Med* **15**, 814-818 (2009).

49. Hanna, J. et al. Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**, 513-524 (2009).

50. Hanna, J.H., Saha, K. & Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143**, 508-525.

51. Rosner, M.H. et al. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686-692 (1990).

52. Scholer, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. & Gruss, P. New type of POU domain in germ line-specific protein Oct-4. *Nature* **344**, 435-439 (1990).

53. Yeom, Y.I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-894 (1996).

54. Pesce, M., Wang, X., Wolgemuth, D.J. & Scholer, H. Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. *Mech Dev* **71**, 89-98 (1998).

55. Niwa, H., Miyazaki, J. & Smith, A.G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-376 (2000).

56. Lenardo, M.J. et al. Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science* **243**, 544-546 (1989).

57. Scholer, H.R., Ciesiolka, T. & Gruss, P. A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells. *Cell* **66**, 291-304 (1991).

58. Liu, L. & Roberts, R.M. Silencing of the gene for the beta subunit of human chorionic gonadotropin by the embryonic transcription factor Oct-3/4. *J Biol Chem* **271**, 16683-16689 (1996).

59. Ben-Shushan, E., Thompson, J.R., Gudas, L.J. & Bergman, Y. Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. *Mol Cell Biol* **18**, 1866-1878 (1998).

60. Botquin, V. et al. New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev* **12**, 2073-2090 (1998).

61. Yamanaka, S. Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cell* **1**, 39-49 (2007).

62. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126-140 (2003).

63. Yuan, H., Corbi, N., Basilico, C. & Dailey, L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**, 2635-2645 (1995).

64. Chew, J.L. et al. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* **25**, 6031-6046 (2005).

65. Okumura-Nakanishi, S., Saito, M., Niwa, H. & Ishikawa, F. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J Biol Chem* **280**, 5307-5317 (2005).

66. Tomioka, M. et al. Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* **30**, 3202-3213 (2002).

67. Dietrich, J.E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134**, 4219-4231 (2007).

68. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-655 (2003).

69. Silva, J. et al. Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737 (2009).

70. Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

71. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-1234 (2007).

72. Schuh, R. et al. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Kruppel, a Drosophila segmentation gene. *Cell* **47**, 1025-1032 (1986).

73. Jiang, J. et al. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* **10**, 353-360 (2008).

74. Ema, M. et al. Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. *Cell Stem Cell* **3**, 555-567 (2008).

75. Nakatake, Y. et al. Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol* **26**, 7772-7782 (2006).

76. Boyer, L.A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-956 (2005).

77. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-440 (2006).

78. Lee, T.I. et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313 (2006).

79. Boyer, L.A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353 (2006).

80. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117 (2008).

81. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S.H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049-1061 (2008).

82. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-635 (2007).

83. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-538 (2006).

84. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567-582 (2008).

85. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**, 118-122 (2009).

86. Hall, J. et al. Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* **5**, 597-609 (2009).

87. Bernstein, E. et al. Dicer is essential for mouse development. *Nat Genet* **35**, 215-217 (2003).

88. Houbaviy, H.B., Dennis, L., Jaenisch, R. & Sharp, P.A. Characterization of a highly variable eutherian microRNA gene. *RNA* **11**, 1245-1257 (2005).

89. Houbaviy, H.B., Murray, M.F. & Sharp, P.A. Embryonic stem cell-specific MicroRNAs. *Dev Cell* **5**, 351-358 (2003).

90. Mineno, J. et al. The expression profile of microRNAs in mouse embryos. *Nucleic Acids Res* **34**, 1765-1771 (2006).

91. Suh, M.R. et al. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol* **270**, 488-498 (2004).

92. Kanellopoulou, C. et al. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19**, 489-501 (2005).

93. Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* **39**, 380-385 (2007).

94. Melton, C., Judson, R.L. & Blelloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621-626.

95. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521-533 (2008).

96. Mak, W. et al. Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303**, 666-669 (2004).

97. Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644-649 (2004).

98. Heard, E. Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol* **16**, 247-255 (2004).

99. Eggan, K. et al. X-Chromosome inactivation in cloned mouse embryos. *Science* **290**, 1578-1581 (2000).

100. Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N. & Lee, J.T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128-132 (2009).

101. Aoto, T., Saitoh, N., Ichimura, T., Niwa, H. & Nakao, M. Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. *Dev Biol* **298**, 354-367 (2006).

102. Efroni, S. et al. Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437-447 (2008).

103. Kobayakawa, S., Miike, K., Nakao, M. & Abe, K. Dynamic changes in the epigenomic state and nuclear organization of differentiating mouse embryonic stem cells. *Genes Cells* **12**, 447-460 (2007).

104. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* **7**, 540-546 (2006).

105. Meshorer, E. et al. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**, 105-116 (2006).

106. Azuara, V. et al. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**, 532-538 (2006).

107. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* **38**, 413-443 (2004).

108. Levine, S.S. et al. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**, 6070-6078 (2002).

109. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326 (2006).

110. Vastenhouw, N.L. et al. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464**, 922-926.

111. Akkers, R.C. et al. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. *Dev Cell* **17**, 425-434 (2009).

112. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88 (2007).

113. Guo, G. et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18**, 675-685.

114. Tang, F. et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6**, 468-478.

115. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-432 (2007).

116. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).

117. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362-365 (1993).

118. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).

119. Meissner, A. et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* **33**, 5868-5877 (2005).

120. Altshuler, D. et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**, 513-516 (2000).

121. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770 (2008).

122. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560 (2007).

123. Surani, M.A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747-762 (2007).

124. Miller, R.A. & Ruddle, F.H. Pluripotent teratocarcinoma-thymus somatic cell hybrids. *Cell* **9**, 45-55 (1976).

125. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr Biol* **11**, 1553-1558 (2001).

126. Cowan, C.A., Atienza, J., Melton, D.A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369-1373 (2005).

127. Tada, M. et al. Pluripotency of reprogrammed somatic genomes in embryonic stem hybrid cells. *Dev Dyn* **227**, 504-510 (2003).

128. Do, J.T. & Scholer, H.R. Nuclei of embryonic stem cells reprogram somatic cells. *Stem Cells* **22**, 941-949 (2004).

129. Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997-1001 (2006).

130. Surani, M.A. Reprogramming a somatic nucleus by trans-modification activity in germ cells. *Semin Cell Dev Biol* **10**, 273-277 (1999).

131. Guan, K. et al. Pluripotency of spermatogonial stem cells from adult mouse testis. *Nature* **440**, 1199-1203 (2006).

132. Kanatsu-Shinohara, M. et al. Generation of pluripotent stem cells from neonatal mouse testis. *Cell* **119**, 1001-1012 (2004).

133. Davis, R.L., Weintraub, H. & Lassar, A.B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000 (1987).

134. Weintraub, H. et al. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc Natl Acad Sci U S A* **86**, 5434-5438 (1989).

135. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

136. Wernig, M. et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-324 (2007).

137. Maherali, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55-70 (2007).

138. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-317 (2007).

139. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol* **25**, 1177-1181 (2007).

140. Boland, M.J. et al. Adult mice generated from induced pluripotent stem cells. *Nature* **461**, 91-94 (2009).

141. Zhao, X.Y. et al. iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86-90 (2009).

142. Kang, L., Wang, J., Zhang, Y., Kou, Z. & Gao, S. iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* **5**, 135-138 (2009).

143. Brambrink, T. et al. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**, 151-159 (2008).

144. Stadtfeld, M., Maherali, N., Breault, D.T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230-240 (2008).

145. Park, I.H. et al. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141-146 (2008).

146. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).

147. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920 (2007).

148. Aoi, T. et al. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321**, 699-702 (2008).

149. Hanna, J. et al. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250-264 (2008).

150. Rideout, W.M., 3rd, Hochedlinger, K., Kyba, M., Daley, G.Q. & Jaenisch, R. Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell* **109**, 17-27 (2002).

151. Hanna, J. et al. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920-1923 (2007).

152. Wernig, M. et al. Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proc Natl Acad Sci U S A* **105**, 5856-5861 (2008).

153. Dimos, J.T. et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218-1221 (2008).

154. Soldner, F. et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964-977 (2009).

155. Park, I.H. et al. Disease-specific induced pluripotent stem cells. *Cell* **134**, 877-886 (2008).

156. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**, 101-106 (2008).

157. Kim, J.B. et al. Direct reprogramming of human neural stem cells by OCT4. *Nature* **461**, 649-643 (2009).

158. Kim, J.B. et al. Oct4-induced pluripotency in adult neural stem cells. *Cell* **136**, 411-419 (2009).

159. Kim, J.B. et al. Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* **454**, 646-650 (2008).

160. Eminli, S., Utikal, J.S., Arnold, K., Jaenisch, R. & Hochedlinger, K. Reprogramming of Neural Progenitor Cells into iPS Cells in the Absence of Exogenous Sox2 Expression. *Stem Cells* (2008).

161. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322**, 945-949 (2008).

162. Yu, J. et al. Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324**, 797-801 (2009).

163. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322**, 949-953 (2008).

164. Woltjen, K. et al. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766-770 (2009).

165. Kaji, K. et al. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* **458**, 771-775 (2009).

166. Kim, D. et al. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell* **4**, 472-476 (2009).

167. Warren, L. et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618-630 (2010).

168. Ichida, J.K. et al. A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell* **5**, 491-503 (2009).

169. Lyssiotis, C.A. et al. Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. *Proc Natl Acad Sci U S A* **106**, 8912-8917 (2009).

170. Maherali, N. & Hochedlinger, K. Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr Biol* **19**, 1718-1723 (2009).

171. Shi, Y. et al. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2**, 525-528 (2008).

172. Ryan, A.K. & Rosenfeld, M.G. POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev* **11**, 1207-1225 (1997).

173. Wernig, M., Meissner, A., Cassady, J.P. & Jaenisch, R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2**, 10-12 (2008).

174. Miura, K. et al. Variation in the safety of induced pluripotent stem cell lines. *Nat Biotechnol* **27**, 743-745 (2009).

175. Feng, B. et al. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* **11**, 197-203 (2009).

176. Heng, J.C., Orlov, Y.L. & Ng, H.H. Transcription Factors for the Modulation of Pluripotency and Reprogramming. *Cold Spring Harb Symp Quant Biol* (2010).

177. Mikkelsen, T.S. et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49-55 (2008).

178. Huangfu, D. et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* **26**, 795-797 (2008).

179. Huangfu, D. et al. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotechnol* **26**, 1269-1275 (2008).

180. Markoulaki, S. et al. Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nat Biotechnol* **27**, 169-171 (2009).

181. Marson, A. et al. Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell Stem Cell* **3**, 132-135 (2008).

182. Silva, J. et al. Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* **6**, e253 (2008).

183. Hanna, J., Carey, B.W. & Jaenisch, R. Reprogramming of somatic cell identity. *Cold Spring Harb Symp Quant Biol* **73**, 147-155 (2008).

184. Wernig, M. et al. A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol* **26**, 916-924 (2008).

185. Hanna, J. et al. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595-601 (2009).

186. Feng, Q. et al. Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells* **28,** 704-712 (2010).

187. Laurent, L.C. et al. Dynamic Changes in the Copy Number of Pluripotency and Cell Proliferation Genes in Human ESCs and iPSCs during Reprogramming and Time in Culture. *Cell Stem Cell* **8,** 106-118 (2010).

188. Chin, M.H. et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5,** 111-123 (2009).

189. Newman, A.M. & Cooper, J.B. Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell* **7,** 258-262 (2010).

190. Guenther, M.G. et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7,** 249-257 (2010).

191. Polo, J.M. et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28,** 848-855 (2010).

192. Kim, K. et al. Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285-290 (2010).

193. Liang, Q., Conte, N., Skarnes, W.C. & Bradley, A. Extensive genomic copy number variation in embryonic stem cells. *Proc Natl Acad Sci U S A* **105,** 17453-17456 (2008).

# Chapter 2

## Reprogramming of murine and human somatic cells using a single polycistronic vector

Bryce W. Carey, Styliani Markoulaki, Jacob Hanna, Kris Saha, Qing Gao, Maisam Mitalipova and Rudolf Jaenisch

This thesis chapter has been published: Carey BW, Markoulaki S, Hanna J, Saha K, Gao Q, Mitalipova M, Jaenisch R. Proc Natl Acad Sci U S A. 2009 Jan 6;106(1):157-62

**ABSTRACT:**

Directed reprogramming of somatic cells by defined factors provides a novel method for the generation of patient-specific stem cells with the potential to bypass both the practical and ethical concerns associated with somatic cell nuclear transfer (SCNT) and human embryonic stem cells (hES). Although the generation of induced pluripotent stem cells (iPS) has proven a robust technology in mouse and human, a major impediment to the use of iPS cells for therapeutic purposes has been the viral-based delivery of the reprogramming factors because multiple proviral integrations pose the danger of insertional mutagenesis. Here we report a novel approach to reduce the number of viruses necessary to reprogram somatic cells by delivering reprogramming factors in a single virus utilizing 2A "self-cleaving" peptides, which support efficient polycistronic expression from a single promoter. We find that up to four reprogramming factors (Oct4, Sox2, Klf4, and c-Myc) can be expressed from a single virus to generate iPS cells in both embryonic and adult somatic mouse cells and we show that a single proviral copy is sufficient to generate iPS cells from mouse embryonic fibroblasts. In addition we have generated hiPS lines from human keratinocytes demonstrating that a single polycistronic virus can reprogram human somatic cells.

## INTRODUCTION

Viral vector mediated transduction of defined factors has been used to generate induced pluripotent stem (iPS) cells from embryonic or adult somatic cells in both mouse and human. By gene expression and developmental potential iPS cells are equivalent to embryonic stem (ES) cells maintaining a full capacity for differentiation with the ability to form teratomas, generate chimeras, and contribute to the germline (1-8). This technology can be readily applied to many cell types in addition to fibroblasts as numerous cell types have been shown to be amenable to direct reprogramming including pancreatic β cells, neural precursors and terminally differentiated B cells (9-11). Without the ethical or practical concerns associated with human embryonic stem cells (hESC), iPS cell technology has emerged as the most promising method for cell-based therapies of regenerative medicine (12, 13). Yet the current protocols involve viral based delivery of reprogramming factors leading to multiple proviral copies throughout the genome in addition to potentially harmful oncogenes, warranting serious concerns regarding insertional mutagenesis or potential re-activation of silenced viral transcripts in cells derived from iPS cell lines.

One improvement to the current strategy would be the delivery of reprogramming factors within the context of a single polycistronic vector. The use of internal ribosomal entry sites (IRES) in polycistronic vectors has been shown to express multiple genes from one promoter but this frequently leads to non-stoichiometric expression and significantly lower levels of the downstream cistron(s). One alternative to this method is autonomous "self-cleaving" 2A peptides originally identified and characterized in apthovirus FMDV (foot-and-mouth disease virus) (14, 15). 2A oligopeptides are generally ~ 18-22 amino acids long and

contain a highly conserved c-terminal D(V/I)EXNPGP motif that mediates 'ribosomal skipping' at the terminal 2A proline and subsequent amino acid "2B" glycine (16). The most well characterized 2A peptides are derived from foot-and-mouth disease virus (FMDV – subsequently referred to as "F2A"), equine rhinitis A virus (ERAV – "E2A"), porcine teschovirus-1 (PTV-1 – "P2A"), and insect *Thosea asigna* virus (TaV – "T2A"). Importantly the ability to express four proteins in vivo has been achieved using 2A peptides (17).

In this work we examined whether polycistronic vectors utilizing 2A peptides can be applied to in vitro reprogramming with the goal of delivering the reprogramming factors in a single polycistronic virus. To test this we developed a four-factor 2A (4F2A) doxycycline (DOX)-inducible lentivirus encoding mouse cDNAs for Oct4, Sox2, Klf4, and c-Myc separated by three different 2A peptides (P2A, T2A, and E2A respectively). Our findings are three-fold: First, we demonstrate that up to three 2A peptides allow expression of the four reprogramming factors in a single vector being sufficient to reprogram both embryonic and adult murine somatic cells. Second, we find that a single proviral copy of the 4F2A is sufficient to reprogram mouse embryonic fibroblasts (MEFs), suggesting a single transgene could potentially reprogram somatic cells. Finally, we show that the 4F2A can be utilized in reprogramming human cells by generating several iPS lines from neonatal foreskin keratinocytes.

## RESULTS:

### GENERATION OF MOUSE iPS CELLS USING A SINGLE POLYCISTRONIC VIRUS

Our goal was to generate polycistronic viral vectors that would express multiple reprogramming genes from a single promoter using 2A peptides. For this one, two, or three 2A oligopeptides containing unique restriction sites were ligated into FUW lentivirus (18) backbones to allow efficient cloning of Oct4, Sox2, c-Myc and Klf4 each separated by a different 2A sequence. Vectors carrying four, three or two factors consecutively with different combinations of F2A, T2A, E2A or P2A sequences (**Fig. 1 *A*)** were tested for their ability to express individual factors by transient transfection in human 293 cells. Western blot analysis demonstrated that 2A peptides support efficient expression of two, three or all four cistrons from a single polycistronic vector (**Fig. 1 *B***).

A tetracycline inducible lentivirus vector was constructed where expression of the genes was controlled from the tetracycline operator minimal promoter (tetOP; **Fig. 1 *C***). To test whether all four genes of a single four-factor (Oct4/Sox2/Klf4/c-Myc) virus could be expressed upon DOX addition MEFs were infected with the polycistronic vector (referred below to as "4F2A") as well as a constitutive FUW lentivirus carrying the tetracycline controllable trans-activator (M2rtTA; abbreviated as rtTA). Two independent experiments were performed and drug inducible expression of the virus was tested 3 days post-infection by qRT-PCR. Using primers for viral specific transcripts (E2A-cMyc), robust induction was observed (7-10 fold) in cells cultured with DOX as compared to control medium (**Fig. 1 *D***). To test the relative induction compared to ES cells, Oct4 and Sox2 primers that cannot discriminate between viral or endogenous transcripts were utilized and in both experiments infected DOX induced MEFs were significantly higher than in ES cells (~ 3.5- and ~17-fold over ES levels respectively). Western blot analysis of cells isolated at 3 days after infection demonstrated that little or no protein was expressed when the cells

were cultured without DOX whereas robust induction was seen in the presence of DOX with levels of Oct4 and Sox2 protein being similar to that in ES cells (**Fig. 1 *E***).

To test whether the 4F2A vector was able to reprogram somatic cells to a pluripotent state MEFs containing a GFP reporter driven by the endogenous Nanog promoter were infected with virus (4F2A+rtTA). 85-90% of the cells stained for Oct4 at 48 hours after transduction indicating high titre infection (**Fig. 2 *A***). Morphological changes were observed a few days after addition of DOX (data not shown) with distinct colonies appearing after about 8 days and Nanog-GFP+ cells at approximately 25 days after DOX induction (**Fig. 2 *B***). After mechanical isolation and subsequent passage the cells had the typical morphology of ES cells and grew independently of DOX. Four independent 4F2A iPS cell lines were established that were positive for the pluripotency markers AP, SSEA1 and Nanog-GFP (**Fig. 2 *C***).

To investigate whether adult somatic cells could be reprogrammed using the 4F2A vector, we infected tail-tip fibroblasts (TTFs) from 14 week-old mice with the 4F2A + rtTA vectors. Similar to MEFs, typical morphological changes were observed a few days after addition of DOX media. Colonies appeared around 8 days and continued to expand until they were picked (day 16) based on morphology. After several passages four stable iPS cell lines were established that stained positive for all pluripotency markers (Nanog, Oct4, SSEA1, AP) (**Fig. 2 *C***). MEF iPS cell lines were injected subcutaneously into SCID mice and were shown to induce teratomas that contained differentiated cells of all three germ layers (**Fig. 3 *A***). Finally, injection of MEF iPS cells (#4) into blastocysts generated postnatal chimeras (**Fig. 3 *B***) demonstrating that a single 4F2A polycistronic virus can reprogram MEFs to a pluripotent state.

To determine the number of proviruses carried in the 4F2A iPS cell lines, DNA was extracted and subjected to Southern blot analysis using an enzyme that does not cut in the vector sequences. Using Oct4, Sox2, c-Myc and Klf4 probes for hybridization, we detected bands of identical molecular weight confirming that the factor sequences were carried in one provirus. The total number of proviruses was between one and three with iPS cell line #4 carrying a single viral insert (**Fig. 3 *C***). One of two integrations from iPS cell line #1 failed to produce a band after c-Myc hybridization, suggesting a 3' deletion of the c-Myc sequences may have occurred. A second digest confirmed the proviral copy numbers (**supporting information (SI) Fig. 1 *A***).

To estimate reprogramming efficiency MEFs were infected with the 4F2A and rtTA vectors and plated at 0.25x10^6 per 10cm plate culture dish. About 70% of the MEFs were infected as estimated by immunostaining of Oct4 at 48 hours after infection (**SI Fig. 2 *A***). Cells were cultured in ES media containing DOX for 20 days and subsequently transferred to ES cell medium until GFP+ colonies were counted on day 25. An average of ~14.7±4 colonies were detected in three independent dishes (10+10+17) indicating a relative efficiency of 0.0001%. This is four-orders of magnitude lower than that of "secondary" fibroblasts or B cells carrying preselected DOX-inducible proviruses (19).

To test the kinetics of reprogramming using the 4F2A virus we performed dox-withdrawl experiments where at specified days (i.e. 2, 4, 8, 12 etc) DOX containing media is replaced with ES media and the number of Nanog-GFP+ colonies are counted at day 25. Using separate drug-inducible viruses to deliver the four factors it has been reported that ~9-12 days is the minimum time required for the generation of stable iPS cells (20, 21). Cells are not passaged during this time in order to

minimize duplication of reprogramming events. Two independent experiments were performed and in both cases single Nanog-GFP+ colonies were present on plates cultured in DOX media for 8 days, similar to the minimum time required using separate viruses (**SI Fig. 1 *B***).

These data demonstrate that a single polycistronic virus containing the four factors linked by three 2A peptides allows factor expression sufficient to generate iPS cells from embryonic or adult somatic cells. Importantly, our results also show that a single polycistronic proviral copy is sufficient to reprogram somatic cells to pluripotency.

## GENERATION OF HUMAN iPS CELLS USING A SINGLE POLYCISTRONIC VIRUS

To investigate whether human cells could be reprogrammed with the polycistronic vector, neonatal human foreskin keratinocytes (NHFK) were transduced with both the constitutive rtTA and DOX-inducible 4F2A vectors. The fraction of infected cells was 10% as determined by staining for Oct4 at 48 hours after transduction (**SI Fig. 3 *A***). Cells were incubated in keratinocyte medium + DOX and allowed to grow for 6 days until they were passaged and cultured in hESC media + DOX on gelatinized plates. Colonies were first detected at day 12 and most displayed transformed morphology with a few colonies exhibiting a distinct appearance that resembled hESC-like morphology. Two such colonies generated in independent infections were picked between 22 and 35 days after infection and found to expand as distinct colonies with morphology similar to hESC (**Fig. 4 *A***). These cells were expanded in the absence of DOX and gave rise to a homogenous population identical to hESC (Ker-iPS) after an additional 2-5 passages. The cells stained for the pluripo-

tency markers AP, Oct4, Nanog, Sox2, SSEA4, Tra1-60, Tra1-81 (**Fig. 4 *B*, SI Fig. 3 *B***) and had a normal karyotype (**Fig. 4 *C***). DNA fingerprinting excluded that such Ker-iPS cell lines were contamination from previously established human iPS cells or hES lines from our lab (data not shown). To determine proviral copy number in Ker-iPS cell lines genomic DNA was extracted and subjected to Southern blot analysis using an enzyme that does not cut in the vector sequences. Probes for all four reprogramming factors show hybridization to similar molecular weight band(s) again indicating they were carried on a single virus. Two different digests (XbaI & BamHI) show the 4F2A proviral copy number is three (#1.1) and two (#3) respectively (**SI Fig. 4 *A-B***).

To test for pluripotency, one line, Ker-iPS #1.1, was injected subcutaneously into SCID mice. These cells induced teratomas and after histological examination differentiated into cells of all three germ layers (**Fig. 4 *D***). In addition, Ker-iPS #1.1 cells, when subjected to an in vitro neural differentiation protocol produced nestin+ neural progenitor cell populations as well as Tuj1+ post-mitotic neurons as detected by immunostaining. (**Fig. 4 *E***).

**DISCUSSION:**

The experiments described in this paper show that up to four different reprogramming factors inserted into a polycistronic vector separated by 2A sequences can be expressed at levels sufficient to achieve reprogramming. This is in contrast to IRES based polycistronic vectors where the down-stream genes are often expressed at substantially lower levels. Embryonic and adult murine fibroblasts as well as postnatal human keratinocytes were induced to form pluripotent iPS cells when infected with the FUW rtTA and 2A vector transducing Oct4, Sox2, Klf4 and c-Myc.

Our finding that a single proviral copy of a polycistronic vector suggests that with sufficient expression it may be possible to reprogram using a stably expressed transgene. This is supported by the reports suggesting insertional mutagenesis is not necessary for reprogramming and that transient expression of reprogramming factors is sufficient to geneate iPS cells (22, 23). Preliminary work by Okita et al. (2008) achieved reprogramming of murine embryonic fibroblasts using Moloney viruses expressing three factors separated by F2A peptides but with a significant reduction in efficiency. We extend these findings by demonstrating that even at a low efficiency 2A vectors can reprogram adult murine somatic cells as well as human keratinocytes.

In addition Okita et al. (2008) were able to reprogram murine embryonic fibroblasts using multiple transient transfections of two plasmids expressing Oct4-Klf4-Sox2 and cMyc respectively. Although iPS cells were generated with greatly reduced efficiency, iPS cell lines with no detectable vector integration were recovered. Interestingly many of the iPS cell lines generated after transient transfection contained integrated plasmids suggesting that stable integration of exogenous sequences can occur.

We observe a reprogramming efficiency significantly lower than previous experiments using single vectors to transduce each of the four factors (**SI Fig. 2 *B***). It is possible that the lower reprogramming efficiency is due to the stochiometry of factor expression from the polycistronic vector, which may be suboptimal for inducing reprogramming. Transduction with separate vectors allows integration of different numbers of proviruses for each factor, therefore reprogramming may select for a specific set of proviral integrations that result in high expression or an optimal stochiometry between the different factors. This may not be possible using the 2A sys-

tem, which has been reported to support near equimolar protein expression *in vivo* (17). Also, when separate vectors transducing each of the four factors were used for induction of iPS cells, Nanog-GFP positive cells were detected as early as 16 days after DOX induction in contrast to GFP positive cells observed 22-25 days after 4F2A vector transduction, consistent with less optimal reprogramming. Moreover, whereas iPS cells frequently carry multiple Oct4 or Klf4 proviruses, consistently fewer Sox2 proviruses were found suggesting that a high level of Sox2 expression may be unfavorable for reprogramming (24).

Ultimately multiple technologies may converge to allow the generation of therapeutically acceptable iPS cells without viral integrations or lingering reprogramming factors within the genome. In a recent proof-of-principle work non-integrating adenoviruses were used alone or in tandem with an inducible Oct4 transgene to reprogram mouse hepatocytes and fibroblasts respectively, demonstrating insertional mutagenesis is not neccessary for reprogramming (23). iPS cells without detectable viral integrations were recovered, however efficiencies were significantly lower than Moloney or lentivirus reprogramming of fibroblasts. It will be important to determine whether other more accessible cell types are amenable to adenovirus-mediated reprogramming as hepatocytes are harder to obtain and grow from human patients in comparison to keratinocytes or fibroblasts. Furthermore transient technologies such as adenovirus or DNA transfection of reprogramming factors may not have the ability to express reprogramming factors throughout the minimum time interval at levels necessary for acquisition of an iPS state, especially important for human reprogramming. Because the four factors are expressed from a defined location the polycistronic vector system described in this paper should sim-

plify the study of reprogramming mechanisms and may facilitate the excision of the vector resulting in iPS cells that carry no exogenous transgenes.

## MATERIALS AND METHODS

### Viral preparation and infection.

Construction of 4F2A lentiviral vectors containing Oct4, Sox2, Klf4, and c-Myc under control of the tetracycline operator and a minimal CMV promoter was generated after EcoRI cloning from a FUW lentivirus backbone. All constructs were generated using unique restriction sites after amplification by PCR to place an individual factor between a respective 2A peptide (1st XbaI-NheI; 2nd SphI; 3rd XhoI; 4th AscI). Respective 2A sequences: P2A – GCCACGAACTTCTCTCTGTTAAAGCAAGCA GGAGATGTTGAAGAAAACCCCGGGCCT; T2A – GAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGA ATCCCGGCCCT; E2A – CAGTGTACTAATTATGCTCTCTTGAAATTGGCTGG AGATGTTGAGAGCAACCCAGGTCCC). Replication-incompetent lentiviral particles (4F2A and M2rtTA) were packaged in 293T cells with a VSV-G coat and used to infect MEFs containing a GFP allele targeted to the endogenous Nanog locus (25) (7). 14-week old tail tip fibroblasts were derived from mice previously published (12). Human keratinocytes (NHFK) were obtained from Coriell Institute for Medical Research Camden, NJ. Viral supernatants from cultures packaging each of the two viruses were pooled, filtered through a 0.45 muM filter and subjected to ultracentrifugation for concentration. Virus pellets were resuspended in ES cell medium (DMEM supplemented with 10% FBS (Hyclone), leukemia inhibitory factor, β-mercaptoethanol (Sigma-Aldrich), penicil-

lin/streptomycin, L-glutamine and nonessential amino acids (all from Invitrogen) before being applied to cells for 24 hours.

**Western Blot**

100 μl of lysis buffer containing 2% SDS, 10 mM dithiothreitol, 10% glycerol, 12% urea, 10 mM Tris-HCl (pH 7.5), 1 mM phenylmethylsulfonyl fluoride, 1x protease inhibitor mixture (Roche), 25 μM MG132 proteosome inhibitor, and boiled for 5 min. Proteins were then quantified using Bradford reagent (Pierce) and taking spectrophotometric readings at 590 nm. Concentrations were estimated against a standard curve generated using bovine serum albumin. Total protein (5 μg) was subjected to electrophoreses in a denaturing 10% polyacrylamide gel containing 10% SDS. Proteins were then transferred onto Immobilon-P membranes (Millipore) using a semi-dry transfer apparatus. Membranes were blocked in PBS, 0.01% Tween 20 containing 2% nonfat powdered milk (Bio-Rad). Proteins were detected by incubating with antibodies at a concentration of 50 ng/ml in blocking solution. Antibodies used were Oct4 (h-134 Santa Cruz Biotechnology); Sox2 (mouse monoclonal R&D Biosystems); c-Myc (06-340 Upstate); Klf4 (H-180 Santa Cruz Biotechnology); GAPDH (sc-25778 Santa Cruz Biotechnology).

**Quantitative RT-PCR**

Total RNA was isolated using Trizol reagent (Invitrogen). Five micrograms of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA Free RNA kit (Zymo Research). One microgram of DNase I–treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately resuspended in 100 mul of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction in an ABI Prism 7000 (Applied Biosys-

tems) with Platinum SYBR green qPCR SuperMix-UDG with ROX (Invitrogen). Equal loading was achieved by amplifying GAPDH mRNA and all reactions were performed in triplicate. Primers used for amplification were as follows:

*Oct4* F, 5'-ACATCGCCAATCAGCTTGG-3' and

R, 5'-AGAACCATACTCGAACCACATCC-3'

*Sox2* F, 5'-ACAGATGCAACCGATGCACC-3' and

R, 5'- TGGAGTTGTACTGCAGGGCG-3'

*4F2A (E2A-cMyc)* F, 5'-GGCTGGAGATGTTGAGAGCAA-3' and

R, 5'-AAAGGAAATCCAGTGGCGC

*GAPDH* F, 5'-TTCACCACCATGGAGAAGGC-3' and

R, 5'-CCCTTTTGGCTCCACCCT-3'

Error bars represent s.d. of the mean of triplicate reactions.

**Southern blotting**

10 μg of BamHI digested genomic DNA was separated on a 0.7% agarose gel, transferred to a nylon membrane (Amersham) and hybridized with $^{32}$P random primer (Stratagene) labeled probes for OCT4 (EcoRI-PstI fragment of pFUW-tetO-OCT4 plasmid), KLF4 (full length KLF4 cDNA), c-MYC (full length c-MYC cDNA) and SOX2 (full length fragment of pFUW-tetO-SOX2 plasmid).

**Immunofluorescent staining**

Cells were fixed in 4% paraformaldehyde for 20 minutes at 25 °C, washed 3 times with PBS and blocked for 15 min with 5% FBS in PBS containing 0.1% Triton-X. After incubation with primary antibodies against Oct4 (Santa Cruz h-134), Sox2 (R&D Biosystems),

Nanog (anti-ms R&D and anti-h), Tra-1-60, (mouse monoclonal, Chemicon International); hNANOG (goat polyclonal R&D Systems); mNANOG (Bethyl A300-398A), Tra1-81 (mouse monoclonal, Chemicon International), SSEA4 and SSEA1 (monoclonal mouse, Developmental Studies Hybridoma Bank) for 1 h in 1% FBS in PBS containing 0.1% Triton-X, cells were washed 3 times with PBS and incubated with fluorophore-labeled appropriate secondary antibodies purchased from Jackson Immunoresearch. Specimens were analyzed on an Olympus Fluorescence microscope and images were acquired with a Zeiss Axiocam camera.

**Mouse Chimera and Teratoma Formation**

Diploid blastocysts (94–98 h after hCG injection) were placed in a drop of Hepes-CZB medium under mineral oil. A flat tip microinjection pipette with an internal diameter of 16 μm was used for iPS cell injections. Each blastocyst received 8-10 iPS cells. After injection, blastocysts were cultured in potassium simplex optimization medium (KSOM) and placed at 37 °C until transferred to recipient females. About 10 injected blastocysts were transferred to each uterine horn of 2.5-day-postcoitum pseudo-pregnant B6D2F1 female. Pups were recovered at day 19.5 and fostered to lactating B6D2F1 mothers when necessary. Teratoma formation was performed by depositing 2x10^6 cells under the flanks of recipient SCID or Rag2-/- mice. Tumors were isolated 3-6 weeks later for histological analysis.

**Human Teratoma formation and analysis**

hiPSCs were collected by collagenase treatment (1.5mg/ml) and separated from feeder cells by subsequent washes with medium and sedimentation of iPSC

colonies. iPSC aggregates were collected by centrifugation and resuspended in a ratio of 10^6 cells in 250µl of iPSC culture media. iPSCs were injected subcutaneously by 21 gauge needle in the back of SCID mice (Taconic). A tumor developed within 6 weeks and the animal was sacrificed before tumor size exceeded 1.5 cm in diameter. Teratomas were isolated after sacrificing the mice and fixed in formalin. After sectioning, teratomas were diagnosed base on hematoxylin and eosin staining. Karyotype analysis was done with CLGenetics (Madison, WI).

**In Vitro differentiation of human IPS cells into neuronal progenitors:**

Human keratinocyte iPS cells were allowed to outgrow in culture without pasaging for 2 weeks with daily medium change. At day 15 after passage distinct neural rossets were observed and picked mechanically by pooled glass pipett (26). Rosettes were replated on dishes precoated with 15µg/ml polyornithin/ 10µg/ml of laminin (Po/Lam) in N2B27 medium supplemented with FGF2 (20ng/ml) EGF (20ng/ml) (All R&D Systems). After 5-7 d cells were dissociated by scraping with cell lifter and pippeting to single cells in N2B27 medium and replated to Po/Lam culture dishes.

**Differentiation and immunocytochemistry**

Induction of differentiation of neural progenitors was performed by withdrowal of FGF2 and EGF from culture medium for 5days. Cells were fixed in 4% paraformaldehyde for 20 min and stained for human nestin (Chemicon; 1:100) and Tuj-1 (1:100) and subsequently washed 3 times with PBS and incubated with fluorophore-labeled appropriate secondary antibodies purchased from Jackson Immunoresearch. Specimens were

analyzed on an Olympus Fluorescence microscope and images were acquired with a Zeiss Axiocam camera.

**Acknowledgements**

We would like to thank D. Hockemeyer and F. Soldner for the FUW M2rtTA (27). M. Wernig, C. Lengner, M. Creyghton, R. Flannery, J. Dausman, D. Fu and members of the Jaenisch lab for excellent assistance and helpful comments. R.J. is supported by grants from the NIH: 5-RO1-HDO45022, 5-R37-CA084198, and 5-RO1-CA087869. J.H. is a Novartis Fellow by the Helen Hay Whitney Foundation.

**REFERENCES**

1. Lowry WE, Richter L, Yachechko R, *et al.* (2008) Generation of human induced pluripotent stem cells from dermal fibroblasts *Proc Natl Acad Sci U S A* 105, 2883-2888.

2. Maherali N, Sridharan R, Xie W, *et al.* (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution *Cell Stem Cell* 1, 55-70.

3. Okita K, Ichisaka T, & Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells *Nature* 448, 313-317.

4. Park IH, Zhao R, West JA, *et al.* (2008) Reprogramming of human somatic cells to pluripotency with defined factors *Nature* 451, 141-146.

5. Takahashi K, Tanabe K, Ohnuki M, *et al.* (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors *Cell* 131, 861-872.

6. Takahashi K & Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors *Cell* 126, 663-676.

7. Wernig M, Meissner A, Foreman R, *et al.* (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state *Nature* 448, 318-324.

8. Yu J, Vodyanik MA, Smuga-Otto K, *et al.* (2007) Induced pluripotent stem cell lines derived from human somatic cells *Science* 318, 1917-1920.

9. Hanna J, Markoulaki S, Schorderet P, *et al.* (2008) Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency *Cell* 133, 250-264.

10. Kim JB, Zaehres H, Wu G, *et al.* (2008) Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors *Nature* 454, 646-650.

11. Stadtfeld M, Brennand K, & Hochedlinger K (2008) Reprogramming of pancreatic beta cells into induced pluripotent stem cells *Curr Biol* 18, 890-894.

12. Hanna J, Wernig M, Markoulaki S, *et al.* (2007) Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin *Science* 318, 1920-1923.

13. Wernig M, Zhao JP, Pruszak J, *et al.* (2008) Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease *Proc Natl Acad Sci U S A* 105, 5856-5861.

14. Ryan MD & Drew J (1994) Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein *Embo J* 13, 928-933.

15. Ryan MD, King AM, & Thomas GP (1991) Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence *J Gen Virol* 72 ( Pt 11), 2727-2732.

16. Doronina VA, Wu C, de Felipe P, *et al.* (2008) Site-specific release of nascent chains from ribosomes at a sense codon *Mol Cell Biol* 28, 4227-4239.

17. Szymczak AL, Workman CJ, Wang Y, *et al.* (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector *Nat Biotechnol* 22, 589-594.

18. Lois C, Hong EJ, Pease S, *et al.* (2002) Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors *Science* 295, 868-872.

19. Wernig M, Lengner CJ, Hanna J, *et al.* (2008) A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types *Nat Biotechnol* 26, 916-924.

20. Brambrink T, Foreman R, Welstead GG, *et al.* (2008) Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells *Cell Stem Cell* 2, 151-159.

21. Stadtfeld M, Maherali N, Breault DT, *et al.* (2008) Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse *Cell Stem Cell* 2, 230-240.

22. Okita K, Nakagawa M, Hyenjong H, *et al.* (2008) Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors *Science* 322, 949-953

23. Stadtfeld M, Nagaya M, Utikal J, *et al.* (2008) Induced Pluripotent Stem Cells Generated Without Viral Integration *Science* 322, 945-949.

24. Eminli S, Utikal JS, Arnold K, *et al.* (2008) Reprogramming of Neural Progenitor Cells into iPS Cells in the Absence of Exogenous Sox2 Expression *Stem Cells.*

25. Meissner A, Wernig M, & Jaenisch R (2007) Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells *Nat Biotechnol* 25, 1177-1181.

26. Zhang SC, Wernig M, Duncan ID, *et al.* (2001) In vitro differentiation of transplantable neural precursors from human embryonic stem cells *Nat Biotechnol* 19, 1129-1133.

27. Hockemeyer D, Soldner F, Cook EG, *et al.* (2008) A drug-inducible system for direct reprogramming of human somatic cells to pluripotency *Cell Stem Cell* 3, 346-353.

## FIGURE LEGENDS

**FIGURE 1: Generation of murine iPS cells using a single 4F2A polycistronic virus.**

A.FUW lentivirus constructs tested by transient transfection. In total four 2A peptides (F2A, T2A, E2A, and P2A) were used.

B.Transient transfection of 293 cells with FUW 2A lentiviruses. Cells were harvested after 48 hours and analyzed by western blot (WB). Efficient protein expression was observed in all constructs tested, indicating four unique 2A peptides support robust protein expression. NOTE: Sox2 protein is not detected in ES cells because only a short exposure was used.

C.Schematic of the 4F2A DOX-inducible lentivirus containing three types of 2A peptides (P2A, T2A, and E2A). Murine cDNAs for Oct4, Sox2, Klf4, and c-Myc. This particular sequence of factors and 2A peptides is subsequently referred to as "4F2A."

D.RT-PCR anaylsis of mRNA induction in cells transduced with OSKM 4F2A+rtTA for 3-days. Total Oct4 or Sox2 induction was used to test levels of 4F2A induction relative to ES cells. E2A-cMyc primers were used to detect viral-specific transcripts. Error bars represent s.d. of the mean of triplicate reactions.

E.Western blot analysis of MEFs transduced with 4F2A+rtTA for three days. Cells infected with 4F2A DOX-inducible lentivirus + rtTA produce all four reprogramming factors upon addition of doxycycline, DOX.

**FIGURE 2: 4F2A iPS cells express pluripotency markers.**

**A.**Immunostaining of Oct4 protein indicates high titre infections can be achieved with the 4F2A. MEFs were cultured in DOX media for 2 days after transduction with 4F2A + rtTA.

**B.**Morphology changes in NanogGFP-MEFs transduced with 4F2A + rtTA cultured in ES media + DOX. Colonies appeared ~ 8 days similar to cells infected with single viruses. Nanog GFP+ colonies were observed by day 25 after DOX media removal at day 20. Two columns show typical colonies observed on the plate.

**C.**4F2A iPS lines generated from Nanog-GFP MEFs or 14-week tail-tip fibroblasts (TTFs) that stain positive for pluripotency markers AP, SSEA1, Oct4 and have reactivated the endogenous Nanog locus (GFP+ for MEFs and by immunostaining for TTF).

**FIGURE 3: 4F2A iPS cells are pluripotent and contain between 1-3 proviral integrations.**

A.In-vivo differentiation of 4F2A MEF-iPS lines #1,2, and 4. Histological analysis of teratomas induced after subcutaneous injection into SCID mice indicates iPS lines contribute to all three germ layers.

B.Moderate to high contribution postnatal chimeric mice as detected by agouti coat color from 4F2A iPS line #4.

Southern blot analysis of 4F2A proviral integrations in MEF-iPS cell lines #1-4. iPS cell DNA was digested with BamHI. Hybridization of the same molecular weight fragment using all four probes indicates presence of 4F2A provirus. Red arrow highlights iPS line #4 which contained one proviral copy of the 4F2A. * indicates endogenous allele.

**FIGURE 4: Generation of human iPS lines using a single 4F2A polycistronic virus.**

A. Neonatal human foreskin keritinocytes (NHFK) transduced with 4F2A (carrying mouse cDNAs) + rtTA. On day 22 a single colony was picked and expanded, giving rise to colonies resembling hES colonies. These colonies were picked and a stable hiPS line was established.

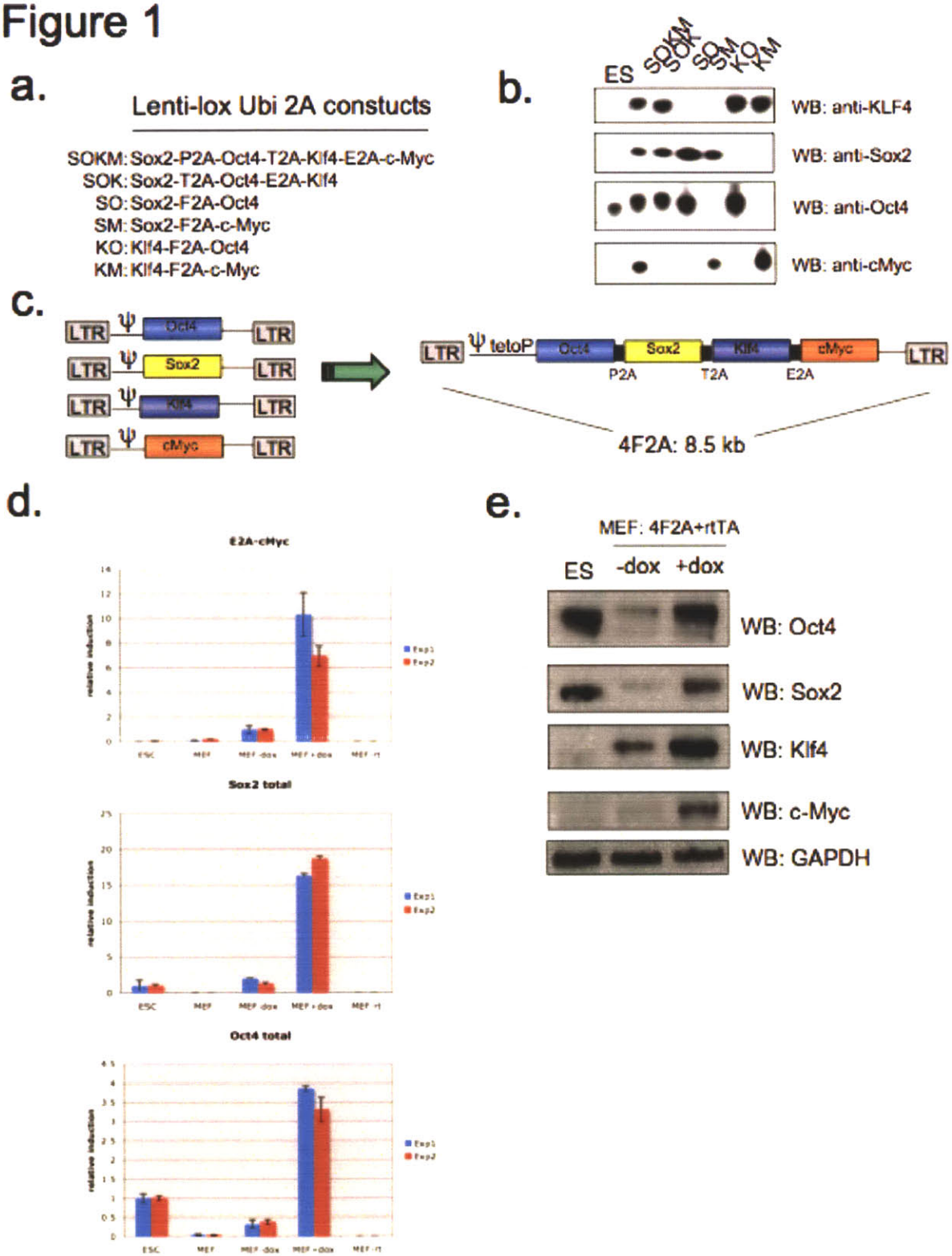
B.Ker hiPS #1.1 immunostaining for pluripotency markers AP, Oct4, Nanog, SSEA-4, Tra1-60, and Tra1-81. DAPI stain is in lower panels.

C.Karyotype of Ker hiPS #1.1 is normal 46 XY.

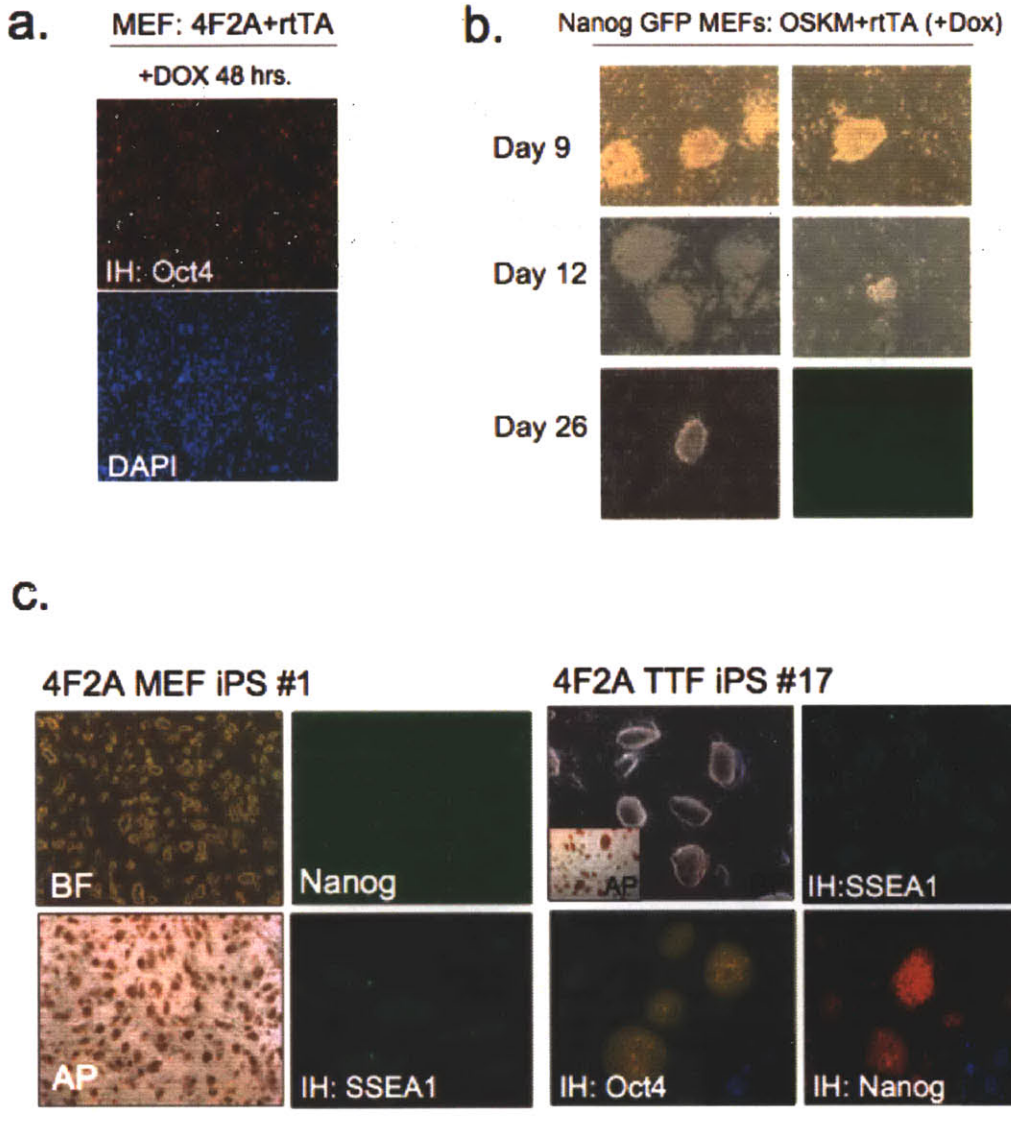
D.*In-vivo* differentiation of Ker hiPS #1.1. Hematoxylin and eosin staining of teratoma sections generated by Ker hiPS #1.1

E.*In vitro* differentiation of Ker hiPS #1.1. (Left) Ker-iPS #1.1-derived neural precursors exposed to differentiation conditions for 6 days produce terminally differentiated neurons as detected by anti-Tuj1 immunostaining (green). (Right) Ker-iPS #1.1 neural precursors (NPs) undergo spontaneous differentiation. NPs were detected by anti-Nestin immunostaining and differentiated neurons by anti-Tuj1 (red). DAPI stain for DNA in both pictures is blue.

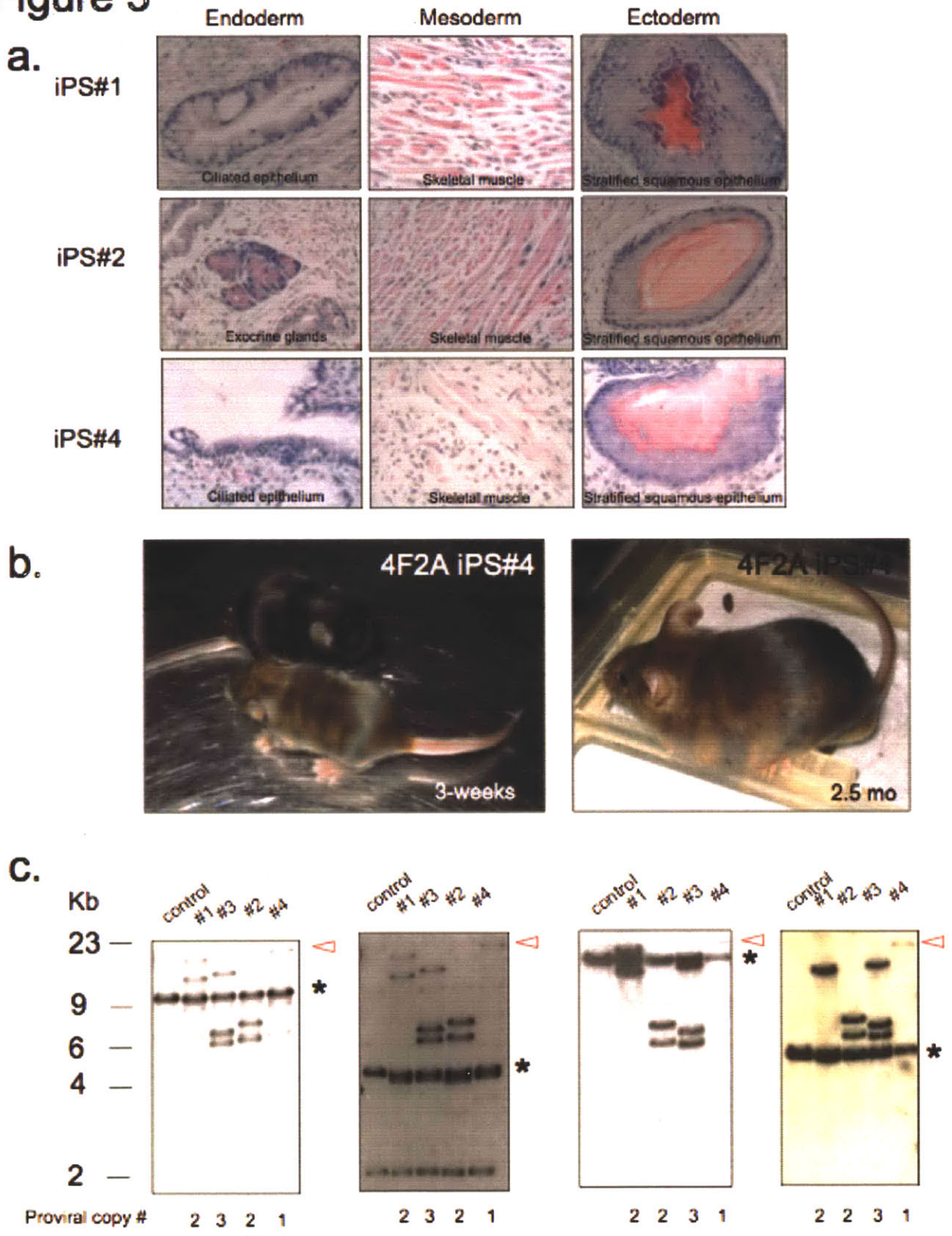
# Figure 1


a.
Lenti-lox Ubi 2A constucts
SOKM: Sox2-P2A-Oct4-T2A-Klf4-E2A-c-Myc
SOK: Sox2-T2A-Oct4-E2A-Klf4
SO: Sox2-F2A-Oct4
SM: Sox2-F2A-c-Myc
KO: Klf4-F2A-Oct4
KM: Klf4-F2A-c-Myc
b.
ES
SOKM
SOK
SO
SM
KO
KM
WB: anti-KLF4
WB: anti-Sox2
WB: anti-Oct4
WB: anti-cMyc
c.
LTR
Ψ
Oct4
Sox2
Klf4
cMyc
tetoP
P2A
T2A
E2A
4F2A: 8.5 kb
d.
E2A-cMyc
Sox2 total
Oct4 total
relative induction
ESC
MEF
MEF -dox
MEF +dox
MEF rt
e.
MEF: 4F2A+rtTA
ES
-dox
+dox
WB: Oct4
WB: Sox2
WB: Klf4
WB: c-Myc
WB: GAPDH

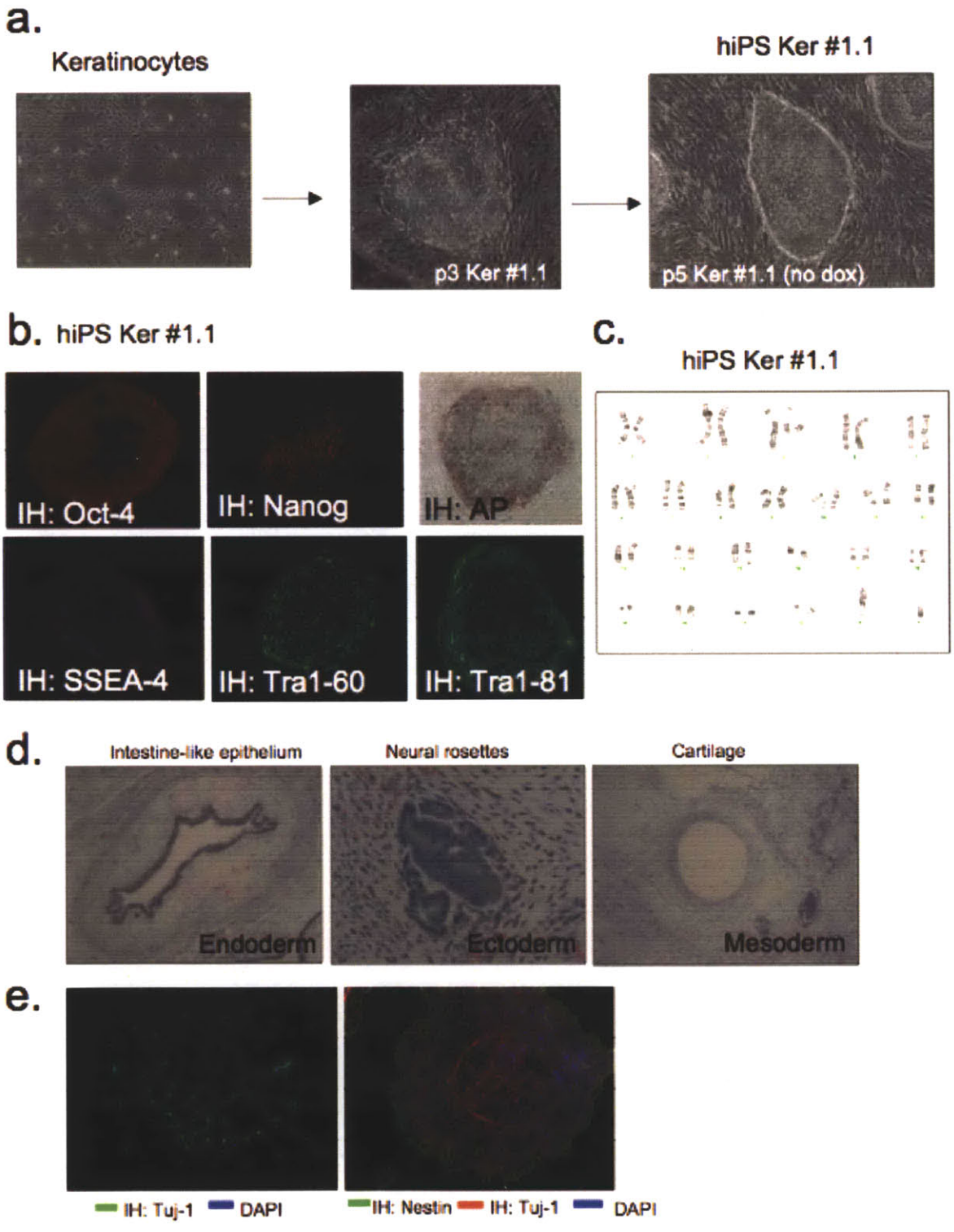
# Figure 2


a.
MEF: 4F2A+rtTA
+DOX 48 hrs.
IH: Oct4
DAPI
b.
Nanog GFP MEFs: OSKM+rtTA (+Dox)
Day 9
Day 12
Day 26
c.
4F2A MEF iPS #1
4F2A TTF iPS #17
BF
Nanog
AP
IH: SSEA1
AP
IH:SSEA1
IH: Oct4
IH: Nanog

# Figure 3


a.
Endoderm
Mesoderm
Ectoderm
iPS#1
iPS#2
iPS#4
Ciliated epithelium
Skeletal muscle
Stratified squamous epithelium
Exocrine glands
Skeletal muscle
Stratified squamous epithelium
Ciliated epithelium
Skeletal muscle
Stratified squamous epithelium
b.
4F2A iPS#4
3-weeks
2.5 mo
c.
Kb
23
9
6
4
2
control
#1 #3 #2 #4
Proviral copy #
2 3 2 1
Probe: Sox2
Probe: Oct4
control
#1 #2 #3 #4
2 2 3 1
Probe: Klf4
Probe: c-Myc

# Figure 4.


a.
Keratinocytes
hiPS Ker #1.1
p3 Ker #1.1
p5 Ker #1.1 (no dox)
b. hiPS Ker #1.1
IH: Oct-4
IH: Nanog
IH: AP
IH: SSEA-4
IH: Tra1-60
IH: Tra1-81
c.
hiPS Ker #1.1
d.
Intestine-like epithelium
Neural rosettes
Cartilage
Endoderm
Ectoderm
Mesoderm
e.
IH: Tuj-1
DAPI
IH: Nestin
IH: Tuj-1
DAPI

**SI FIGURE 1: Southern blot of MEF-derived iPS lines and dox-withdrawl indicating 8 days is sufficient to generate iPS lines.**

A.Southern blot analysis of 4F2A MEF iPS lines. A second digest was performed (XbaI) to confirm the proviral copy number. In this digest iPS line #2 and #4 show 1 proviral copy, however only #4 had 1 proviral copy in both digests.

B.Dox-withdrawl after 8 days post-infection of Nanog GFP MEFs with rtTA+OSKM generated two iPS lines. Both generated stable iPS lines after 1-2 passages.

**SI FIGURE 2: Relative efficiencies reprogramming using 4F2A**

A.Relative reprogramming efficiency of 4F2A in MEFs. NanogGFP MEFs were infected with 4F2A+rtTA and cultured in ES media (+/- DOX) for 48 hours. Cells were fixed and stained for Oct4 protein. An estimated infection efficiency was ~ 70%. The same virus was also used to infect 0.2 5x10^6 Nanog GFP MEFs and cells were cultured on DOX for 20 days. After withdrawl of DOX at day 20, GFP+ colonies were counted at day 25, in three plates 10, 10, and 17 GFP+ colonies were observed.

**SI FIGURE 3: Infection efficiency and pluripotency analysis of Keratinocyte-derived human iPS lines.**

A.Infection efficiency from two experiments as detected by Oct4 immunostaining in Keratinocytes infected with 4F2A+rtTA and cultured in hES media+DOX for 48 hours. Efficiency of infection was ~10-20% based on fraction of cells positive for Oct4 protein.

B.Human iPS lines stain positive for pluripotency markers expressed in hES cells (Ker iPS #3 is shown).

**SI FIGURE 4: Proviral copy number of Keratinocyte-derived human iPS lines.**

**A.** Southern blot analysis of Ker-iPS lines. 10mg of genomic DNA was harvested and digested with XbaI. Hybridization of the same molecular weight fragment indicates presence of 4F2A provirus. Probes for Sox2, Klf4, and c-Myc suggested 2 (#1.1) and 1 (#3) proviral copies. Common bands observed between the two iPS lines are not viral integration as these were derived from independent infections.

B. Southern blot analysis of Ker-iPS lines. 10mg of genomic DNA was harvested and digested with BamHI. Hybridization of the same molecular weight fragment indicates presence of 4F2A provirus. Probes for Oct4 and c-Myc indicate 3 (#1.1) and 2 (#3) proviral copies.

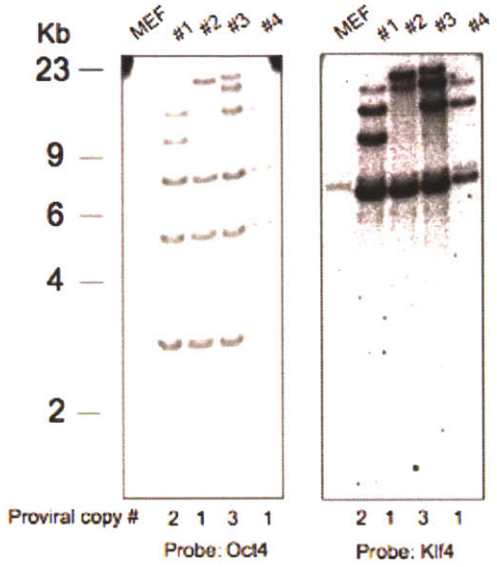
**Table S1:** Table summarizing pluripotency tests as well as relative efficiencies for all iPS lines generated. GFP, GFP reporter gene present; ES, expression of ES cell markers (AP, SSEA1, Oct4 or Sox2); TF, teratoma formation; PC, postnatal chimeras. Mouse chimerism was estimated by agouti coat color.

| Source of cells | GFP | iPS lines | Efficiency (iPS/input, %) | ES | TF | PC |
|---|---|---|---|---|---|---|
| (m) embryonic fib | Nanog | 5 | 0.0001% | Yes | Yes | Yes |
| (m) adult fib | No | 4 | ND | Yes | No | No |
| (h) Keratinocytes | No | 2 | 0.00001% | Yes | Yes | No |

| Cell line | Blast injected | Live pups | # chimeric | chimerism (%) |
|---|---|---|---|---|
| MEF iPS #4 | 60 | 30 | 2 | 30-50 |
| MEF iPS #2 | 20 | 14 | 1 | 10 |

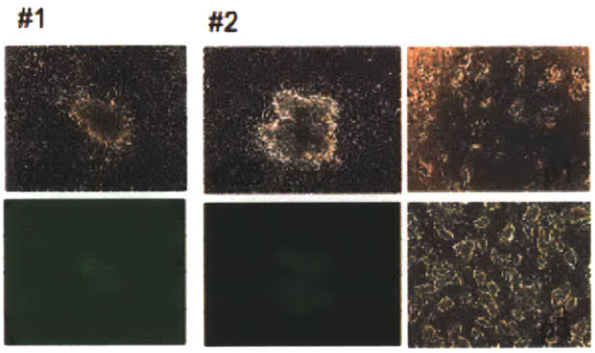
# SI Figure 1

**a.** XbaI digest

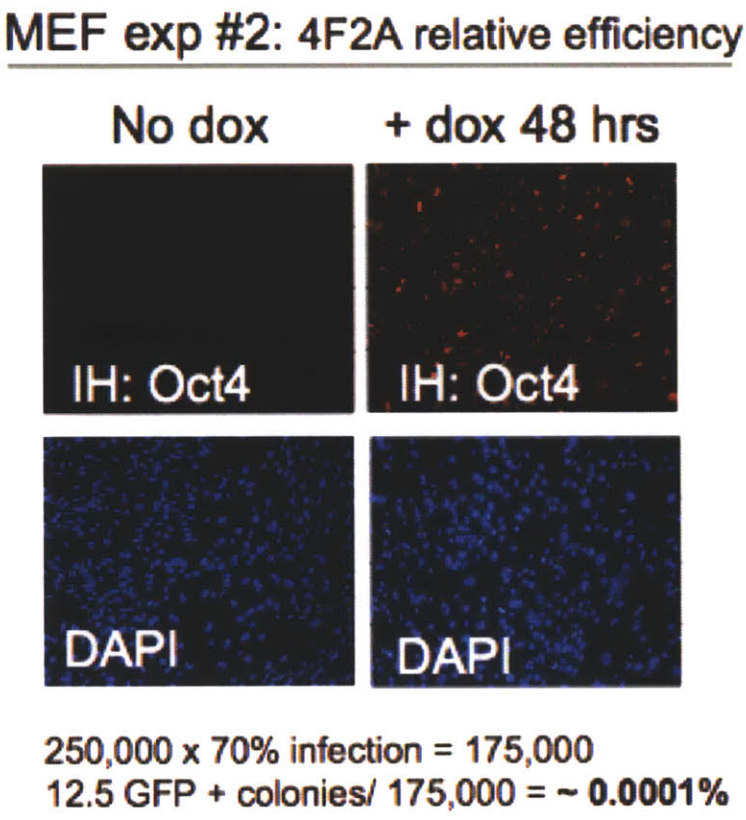



**b.**

8 days of DOX exposure

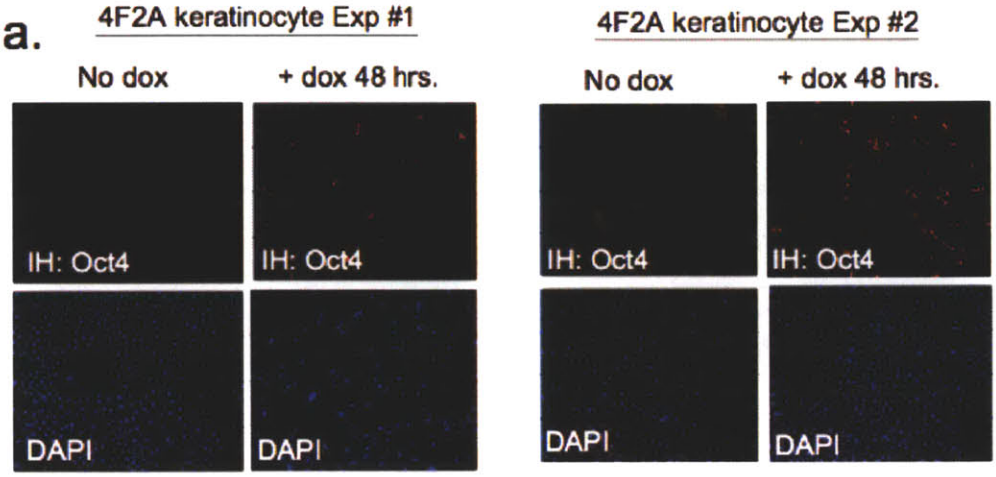


# SI Figure 2

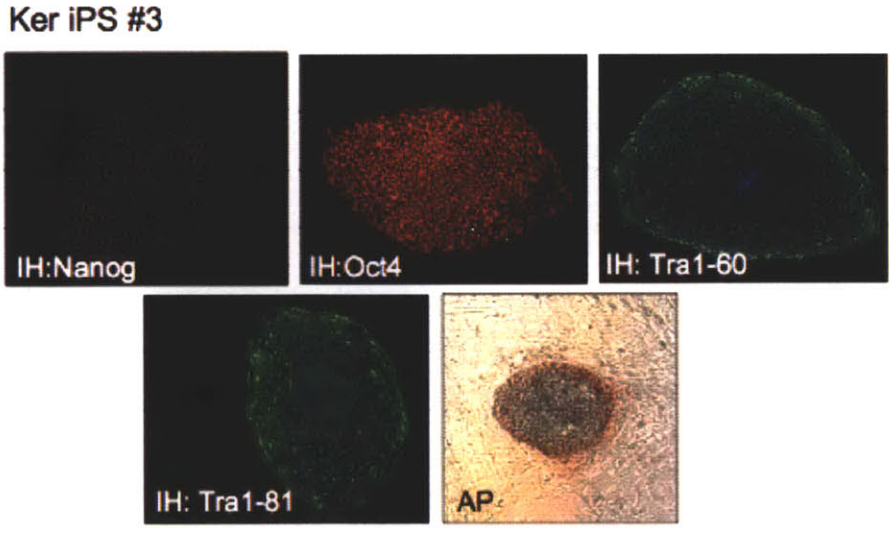

MEF exp #2: 4F2A relative efficiency
No dox
+ dox 48 hrs
IH: Oct4
IH: Oct4
DAPI
DAPI
250,000 x 70% infection = 175,000
12.5 GFP + colonies/ 175,000 = ~ 0.0001%

# SI Figure 3

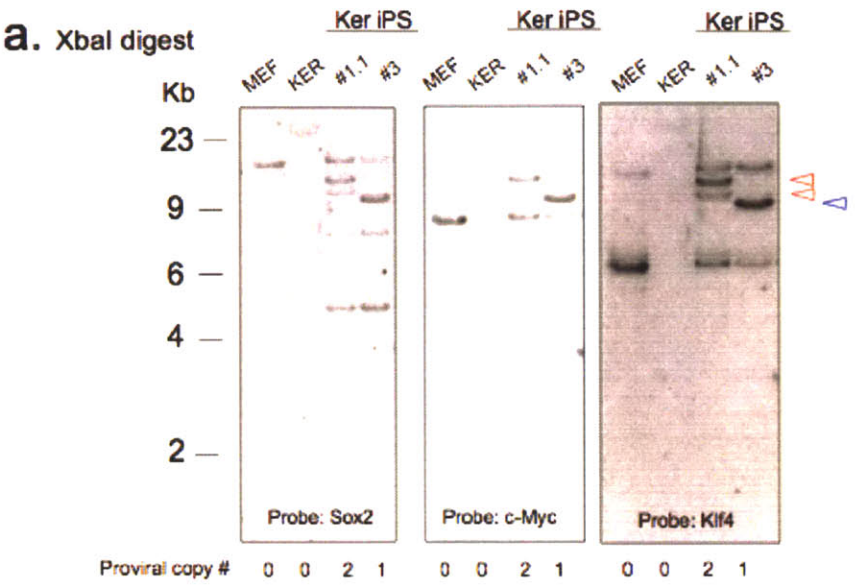
a.

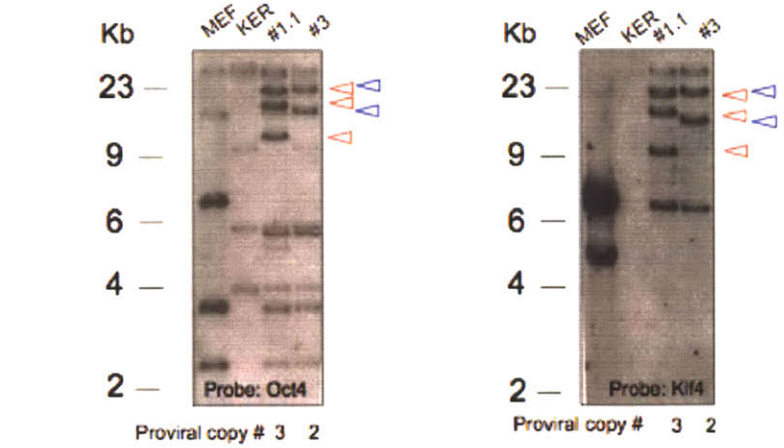



b.



# SI Figure 4


a. XbaI digest
Ker iPS
Ker iPS
Ker iPS
MEF
KER
#1.1
#3
Kb
23
9
6
4
2
Probe: Sox2
Probe: c-Myc
Probe: Klf4
Proviral copy # 0 0 2 1
0 0 2 1
0 0 2 1


b. BamHI digest
Kb
MEF
KER
#1.1
#3
23
9
6
4
2
Probe: Oct4
Proviral copy # 3 2
Probe: Klf4
Proviral copy # 3 2

# Chapter 3

## A single-gene transgenic mouse strain for reprogramming adult somatic cells

Bryce W. Carey, Styliani Markoulaki, Caroline Beard, Jacob Hanna, and Rudolf Jaenisch

This thesis chapter has been published: Carey BW, Markoulaki S, Beard C, Hanna J, Jaenisch R. *Nat Methods*. 2010 Jan;7(1):56-9.

**ABSTRACT: We report a novel transgenic mouse model for direct reprogramming of somatic cells by expressing four reprogramming factors from a single genomic locus using a drug-inducible transgene. Multiple somatic cell types explanted from different tissues can generate induced pluripotent stem cells (iPSC) by culture in doxycycline (Dox). Because the reprogramming factors are carried on a single polycistronic construct the transgenic mice can be easily maintained and transferred into another background.**

**RESULTS**: Generation of induced pluripotent stem cells (iPSCs) by primary infection using Moloney based gene delivery generates iPSCs with an efficiency between 0.001-0.1% [1-4]. The creation of "secondary" systems that utilize doxycyline (Dox)-inducible lentiviruses to generate iPSCs from murine somatic cells allows the reprogramming of genetically homogenous somatic cell populations and avoids the genetic heterogeneity of primary infections [5-8]. However secondary lines harbor multiple (>7) proviruses, each with different capacities for reactivation as indicated by the variegated expression of reprogramming factors after Dox exposure in addition to the inability of factors to reactivate in some somatic tissues [6]. Although secondary models have been used to generate transgenic animals carrying all or a subset of factors by germline segregation of the proviruses, it remains difficult to maintain all factors in one mouse because the multiple transgenes segregate in each generation [9].

Here we report a transgenic mouse model that overcomes these limitations and minimizes the variegation of factor levels by inducing reprogramming factor expression

from a single genomic site. The collagen type I (*Col1a1*) gene has been used as an expression locus for *Oct4* (also called *Oct3/4* or *Pou5f1*) and demonstrated to be capable of transgene induction in multiple somatic cell types including mouse fibroblasts, keratinocytes, adrenal glands, and neural precursor cells [6, 10, 11]. We generated "single-gene" transgenic mouse strain(s) that express three or four reprogramming factors from the *Col1a1* locus.

Using homologous recombination of embryonic stem cells (ESCs) carrying a drug-inducible reverse tetracycline trans-activator (M2rtTA) in the Rosa26 locus, we inserted four reprogramming factors as a single polycistronic transgene in an expression cassette termed "4F2A," which harbors four murine reprogramming genes (*Oct4*, *Sox2*, *Klf4*, and *c-Myc*) separated by three 2A self-cleaving peptides (**Fig. 1a,b**). This cassette has been shown to reprogram embryonic and adult murine fibroblasts as well as human keratinocytes [12]. Transgenic animals carrying one copy of each transgene rtTA(1):4F2A(1) were generated and intercrossed to derive animals carrying various combinations of transgene copies rtTA(2): 4F2A(1), rtTA(1):4F2A(2) and rtTA(2):4F2A(2). Importantly, no tumors or other health problems were detected in any of the transgenic mice for up to 31 weeks of age (**Supplementary Table 1**).

We used qRT-PCR analysis to quantify the Dox-induced expression level of the 4F2A transgene (Tg) in mouse embryonic fibroblasts (MEFs) using primers specific to the 4F2A Tg and observed >100-fold induction in MEFs carrying a single copy of both the rtTA and 4F2A (**Supplementary Figure 2a**). Protein expression for three of the four factors was confirmed by immunoflourescence demonstrating robust Dox-inducible expression (**Supplementary Figure 2b**). The relative induction as compared to transcript levels in ESCs or MEFs was quantified by qRT-PCR analysis using exon specific primers for each factor demonstrating that the expression of *Oct4* and *Sox2* was in the same

range whereas *Klf4* and *c-Myc* were expressed at a 3- to 5-fold higher than in ES cells (**Supplementary Fig. 3a**). To test whether factor expression was copy number dependent we quantified transcript levels in adult fibroblasts carrying different copies of the two transgenes. qRT-PCR analysis showed a copy number and Dox-dependent expression of the 4F2A Tg (1:1 < 1:2 < 2:2; **Supplementary Fig. 3b**).

Mice carrying different copies of each transgene (1:1, 2:1, 1:2, 2:2) were tested for the ability to generate iPSCs from defined tissues (**Supplementary Table 2**). MEFs isolated from animals carrying single copy rtTA(1):4F2A(1) and cultured in the presence of Dox gave rise to colonies within 2 – 4 weeks that could be picked and after 2-3 passages generated iPS cell lines that expressed the pluripotency markers alkaline phosphatase (AP), SSEA1, and had reactivated the endogenous *Nanog* locus as detected by immunostaining (**Fig. 1c**). However, we failed to obtain iPSCs from any tissue of adult mice carrying single copies of each Tg respectively.

To assess whether higher transgene expression levels would allow the generation of iPSCs from other cell types, tail tip fibroblasts (TTF), epidermis-derived keratinocytes, liver cells, mesenchymal stem cells (MSCs), pro-B cells, macrophages and intestinal epithelial cells were derived from mice carrying different combinations of Tg copies and cultured in the presence of Dox. We were able to isolate iPSCs from adult liver cells as well as keratinocytes from mice carrying 2 copies of rtTA and 1 copy of 4F2A (rtTA(2): 4F2A(1)). Colonies appeared at d12-16 after addition of Dox (**Supplementary Figure 4a**), were manually picked, expanded without Dox after 2-3 passages and shown to express pluripotency markers AP, SSEA1, and Nanog (**Fig. 1c**). Other cell types such as CD11b+ macrophages from spleen, CD19+ pro-B cells from bone marrow, intestinal epithelium, adult tail-tip fibroblasts, and mesenchymal stem cells (MSCs) generated iPSCs only in the presence of a second copy of 4F2A transgene. iPSCs from adult fibro-

blasts, CD11b+ macrophages, and pro-B cells were derived from animals carrying 2 copies of the 4F2A Tg and one copy of the rtTA Tg (rtTA(1):4F2A(2)) whereas intestinal epithelial cells and mesenchymal stem cells (MSCs) required the presence of a second rtTA Tg (rtTA(2):4F2A(2)). The iPSCs grew in the absence of Dox and expressed the pluripotency markers AP, SSEA1, and Nanog (**Fig. 1c** and **Supplementary Figure 4b**). Complete demethylation of the endogenous *Oct4* promoter was observed in five iPSC lines tested after bisulfite treatment and sequencing (**Fig. 1d** and **Supplementary Figure 4c**).

To assess pluripotency we tested the ability of Col1a1 4F2A iPSCs to differentiate *in vivo*. Liver, keratinocyte, and MEF-derived iPSCs were injected sub-cutaneously into SCID mice, giving rise to teratomas containing cells from all three germ layers (**Figure 1e** and **Supplementary Fig. 4d**). Two lines were tested for the ability to contribute to post-natal chimeras. MEF iPSC line #11 contributed to adult chimeras after injection into blastocysts (**Fig. 1f**). In addition we generated high contribution post-natal chimeras from macrophage-derived iPSC line (#5) that also transmitted the iPSC genome through the germline (**Fig. 1f** and **Supplementary Figure 4e**).

Our data demonstrate that iPSCs can be derived from multiple somatic tissues of adult mice carrying the Col1a1 4F2A polycistronic transgene. The observations that some donor cell types were able to generate iPSCs only from animals with two copies of the transgene suggests cell types such as intestinal epithelial cells, MCSs, pro-B cells or TTFs require higher levels of factor expression for reprogramming than other cell types such as MEFs or hepatocytes. Alternatively, differences in the induction level of the Col1a1 locus in certain cell types may require multiple copies of a Col1a1 transgene to reach a threshold of expression capable of achieving an iPS cell state.

We characterized reprogramming efficiencies in two distinct mouse transgene ratios (1:2 and 2:2) using pro-B cells. From each strain CD19+ cells were harvested from bone marrow and plated at similar densities ($1x10^6$) per well. Colonies appeared in both genotypes around day 12-14 but with different efficiencies. Alkaline phosphatase (AP) and Nanog immunostaining showed a 100-fold higher efficiency at a 2:2 than at a 1:2 transgene ratio (**Supplementary Fig. 5a,b**). To quantify the efficiency more precisely we tested the kinetics of reprogramming CD19+ pro-B cells (using ratio 2:2). Approximately $5x10^5$ cells were cultured with Dox and changed to ES media without dox at indicated times. Reprogramming was assessed and quantified as Dox-independent, Nanog-positive iPS cells by immunostaining on day 25. We found Nanog-positive colonies increased with prolonged culture and directly correlated with the number of input donor cells (**Supplementary Figure 5c,d**).

In addition to the Col1a1 4F2A system we established a transgenic model carrying only a subset of reprogramming factors to demonstrate the potential utility for replacement of individual reprogramming factors[9]. Using a three-factor 2A expression cassette lacking *Sox2* (Oct4-T2A-Klf4-E2A-cMyc; abbreviated as "3F2A") western blot analysis demonstrated that cells transduced with the 3F2A lentivirus were capable of expressing all three reprogramming factors (**Fig. 2a**). This cDNA was used in a Col1a1 targeting vector to generate Col1a1 3F2A transgenic ESCs that also harbored a drug-inducible M2rtTA at the Rosa 26 locus (**Fig. 2b**). After injection of 3F2A ESCs into blastocysts we obtained E13.5 chimeric embryos and isolated MEFs. When the cells were cultured in the presence of Dox induction of the transgene was ~ 120-fold as seen by qRT-PCR (**Supplementary Fig. 6a**). Because the 3F2A-MEFs lacked the essential reprogramming factor Sox2 they were infected with a Sox2 lentivirus (pFUW-tetO Sox2) (**Fig 2c**). Colonies appeared after 12-16 days and were picked at day 25. iPSCs were estab-

lished after 2-3 passages staining positive for pluripotency markers AP, SSEA1 and Nanog (**Fig. 2d**). Four independent iPSC lines (MEF 3F2A+S iPSC) carried a single Sox2 provirus as detected by Southern (**Fig. 2e**) in addition to the single copy of rtTA and OKM Tg. All MEF 3F2A+S iPSC lines carried the 3F2A Tg as demonstrated by Southern analysis using probes for Oct4, Klf4, and c-Myc which all hybridized to a ~4.8 kb band as expected for the single 3F2A transgene (**Supplementary Fig. 6b,c**). iPSC line #34 gave rise to high contribution chimeras when injected into blastocysts (**Fig 2f** and **Supplementary Table 3**) with two of the four chimeras transmitting the iPSC genome through the germline (**Fig. 2f**). Somatic cells derived from animals carrying the 3F2A transgene will be useful to screen for small molecules that replace Sox2 for reprogramming.

Recently, several strategies have been described that allow the generation of vector-free iPSCs following excision of reprogramming transgenes [13-15]. To establish a system that would permit the isolation of vector-free iPSCs from a defined transgenic strain, we created a Col1a1 4F2A targeting vector harboring loxP sites flanking the Tg (Col1a1 2lox 4F2A) (**Fig. 3a**) leaving a single loxP site in the *Col1a1* 3′ UTR after Cre-mediated excision. The construct was targeted to the *Col1a1* locus in ESCs carrying the M2rtTA transgene at the Rosa 26 locus (**Fig 3b**). To assess whether the transgene could be excised freshly passaged transgenic ESCs were exposed to a recombinant cell-permeable Cre protein (HTN-Cre) for 20 hours. Deletion of the Col1a1 2lox 4F2A was detected in sibling colonies that were cultured in the presence or absence of G418 giving rise to 2 out of 16 neo sensitive colonies (**Fig. 3c**) that had deleted the transgene as verified by Southern blot analysis using an external *Col1a1* 5′ probe as well as probes for Oct4 and c-Myc cDNA (**Fig. 3d**). Mice derived from this ES cell line will be useful as donors for generating iPSCs that carry no vector sequences.

The transgenic models described in this study have several advantages over previous systems: 1) the previous "secondary" iPSCs were derived from primary iPS cells generated by lentivirus-mediated gene transfer of the four reprogramming factors [6,9]. Because the individual proviruses carrying the four factors segregate independently it has been difficult to derive mice homozygous for all factors that could be used as a convenient source for reprogrammable cells. In the present system all reprogramming factors are expressed from one cassette (e.g. 4F2A) inserted into a single locus, facilitating the maintenance of the strain as well as the transfer of the reprogramming factors to mouse strains carrying other genetic determinants of interest. 2) Unlike proviral transgenes that may be silenced by DNA methylation or other epigenetic mechanisms upon germ line transmission, this transgenic system expresses the four reprogramming factors from a specific locus (*Col1a1*) that is widely expressed in the adult mouse and is capable of reprogramming cells from seven different tissues of the adult. 3) No tumors or other adverse effects have been observed in the transgenic mice suggesting that the Col1a1 4F2A system is both safe and practical for extended breeding. 4) Reprogramming factor stochiometries are fixed within each somatic cell type. This system creates iPSCs from multiple somatic tissues and allows comparison among iPSCs not previously possible. Moreover this system generates iPSCs that can now be compared to genetically identical ESCs. 5) The generation of 3F2A transgenic mice lacking one of the 4 factors will allow screening of somatic cells for small molecules to replace the fourth factor (such as Sox2, **Fig. 2**). 6) Finally, the generation of mice carrying the loxP flanked 4F2A construct will allow the excision of the transgene and derivation of adult mice created from vector-free iPSCs [13-15].

## MATERIALS AND METHODS:

**Col1a1 targeting and vectors** (1) Col1a1 4F2A: A ~ 5.8kb EcoRI generated fragment containing Oct4-P2A-Sox2-T2A-Klf4-E2A-c-Myc (also named "4F2A") cDNA was generated from a previously published plasmid FUW-Ubi-4F2A [12] and ligated into ptet.splicePL3 vector (containing tetO promoter + SV40 intron,pA) similarly digested with EcoRI to obtain ptet.splicePL3-OSKMpA. A 7.3kb NotI generated fragment was obtained from the ptet.splicePL3-OSKMpA and ligated into a NotI digested mCol.loxneo targeting vector that contained both 5' and 3' homology arms towards the 3' UTR of the Col1a1 locus as well as a pgk-driven neo resistance cassette for selection of transgenic cells. The resulting ~18 kb targeting construct (Col1a1 4F2A) was linearized with PvuI restriction enzyme digestion (25 μg), precipitated and resuspended in 400μl HEPES, then electroporated into V19 ES cells (V6.5 ES cells containing a reverse tetracycline trans-activator (M2rtTA) driven by the Rosa26 promoter). After 24 hours G418 (350 μg per ml) was added to ES cell medium (DMEM supplemented with 10% FBS (Hyclone), leukemia inhibitory factor, β-mercaptoethanol (Sigma-Aldrich), penicillin/streptomycin, L-glutamine and nonessential amino acids (all from Invitrogen). Neo-resistant colonies were picked 10 days later, expanded and DNA isolated for testing of correct targeting by Southen analysis. (2) Col1a1 2lox 4F2A: the 7.3 kb NotI generated fragment from ptet.splicePL3-OSKMpA was blunt-ended and ligated with a EcoRV digested pBS246 plasmid (containing two loxP sites). The resulting pBS246.tetO-OSKMpA was digested with NotI generating a fragment of 7.6 kb that was ligated into the previously mentioned NotI digested mCol.loxneo targeting vector. The resulting vector Col1a1 2lox 4F2A was digested for targeting of V19 ES cells as described above. (3) Col1a1 3F2A: A three-factor cDNA containing Oct4-T2A-Klf4-c-Myc was generated by

PCR amplification of Oct4 using 5' XbaI-EcoRI and SphI containing primers whose product was then TOPO cloned following manufactures instructions (pCR 2.1 TOPO Invitrogen). The TOPO-Oct4 plasmid was then digested with XbaI and SphI and the 1.1 kb Oct4 cDNA was ligated to a similarly digested FUW-Ubi-4F2A ~ 10 kb fragment. The resulting FUW-Ubi-3F2A was digested with EcoRI to generate a ~ 4.0 kb fragment that was cloned into the ptet.splicePL3 vector as described above. A 6.3 kb fragment from ptet.splicePL3-tetO-3F2ApA was generated by NotI digestion and ligated into the mCol.loxneo targeting vector. The resulting 17kb plasmid (Col1a1 3F2A) was linearized with PvuI restriction enzyme (25 μg), precipitated, resuspended in 400 μl HEPES and electroporated into V19 cells. Selection was the same as described above.

**Mouse chimera and teratoma formation:** Diploid blastocysts (94–98 h after hCG injection) were placed in a drop of Hepes-CZB medium under mineral oil. A flat tip microinjection pipette with an internal diameter of 16 μm was used for iPS cell injections. Each blastocyst received 8-10 iPS cells or ES cells. After injection, blastocysts were cultured in potassium simplex optimization medium (KSOM) and placed at 37 °C until transferred to recipient females. About 10 injected blastocysts were transferred to each uterine horn of 2.5-day-postcoitum pseudo-pregnant B6D2F1 female. Pups were recovered at day 19.5 and fostered to lactating B6D2F1 mothers when necessary. For summary see **Supplementary Table 4**. Teratoma formation was performed by depositing 2x10^6 cells under the flanks of recipient SCID mice. Tumors were isolated 3-6 weeks later for histological analysis.

**Somatic cell isolation and culture:** For MEF isolation, chimeric embryos were isolated

at E13.5, and the head and internal organs were removed. The remaining tissues were physically dissociated and incubated in trypsin at 37°C for 20 min after which cells were resuspended in MEF medium. 24 hours later puromycin (2 μg/ml) was added and the cells were expanded for two passages before freezing or plating in doxycline containing media (2 μg/ml) for reprogramming experiments. Somatic organs were isolated from 4-6 week-old transgenic mice. Epidermal keratinocytes and intestinal epithelium were isolated and cultured as previously described [6]. For mesenchymal stem cells (MSCs) and pro-B cells whole marrow was isolated from the femur and tibia after removal of the condyles at the growth plate by flushing with a syringe and 30-guague needle containing DMEM+5% FBS (Hyclone, Thermo Fisher Scientific). CD19+ pro-B cells were isolated by MACS cell separation (Miltenyibiotec Cat# 130-052-201) following manufactures instructions. Purified B cell subsets were resuspended in IMDM with 15% FCS as well as IL-4, IL-7, SCF (10 ng/ml each, Peprotech), doxycyline (2 μg/ml) and plated on OP9 bone marrow stromal cells (ATCC). Three days later the medium was changed to ESC medium plus Dox. Macrophage cells (CD11b+) from freshly isolated spleen were isolated by MACS cell separation and plated by a similar protocol described for CD19+ pro-B cells. Mesenchymal stem cells were selected through differential plating on tissue culture plates (10cm) for 72h in α-MEM supplemented with 15% FBS. Once plates reached a full monolayer cells were split into 6-well dishes and cultured in the presence of Dox. For isolation of liver cells mice were first perfused with 50 ml HBSS buffer (w/o Ca2+ and Mg2+) then 50 ml HBSS (w/o Ca2+ and Mg2+) containing collagenase (type IV) (Sigma Cat# C5138) (100U/ml). Liver was dissected away from surrounding tissues and dissociated in 10ml DAG media (phenol-red free EMEM Gibco-11054-020 and Bovine serum albumin (BSA) 1g/0.5L) and filtered two times through a sterile 100 μM cell

strainer. Liver cell preparations were centrifuged at 30 g for 3 minutes at 4 °C and the cells were washed two times with DAG media and then plated on γ-irradiated MEFs in ES media + Dox.

**Antibodies and Immunostaining:** Cells were fixed in 4% paraformaldehyde for 20 minutes at 25 °C, washed 3 times with PBS and blocked for 15 min with 5% FBS in PBS containing 0.1% Triton-X. After incubation with primary antibodies against Oct4 (Santa Cruz H-134), Sox2 (R&D Biosystems), Klf4 (Santa Cruz H-180), Nanog (Bethyl A300-398A), SSEA1 (monoclonal mouse, Developmental Studies Hybridoma Bank) for 1 h in 1% FBS in PBS containing 0.1% Triton-X, cells were washed 3 times with PBS and incubated with fluorophore-labeled appropriate secondary antibodies purchased from Jackson Immunoresearch. Specimens were analyzed on an Olympus Fluorescence microscope and images were acquired with a Zeiss Axiocam camera. Alkaline phosphatase (AP) staining was done following manufactures instructions (Vector Labs).

**Bisulfite Sequencing** DNA was isolated from indicated cell lines and conversion of unmethylated CpG dinucleotides was performed using EpiTect bisulfite kit (Qiagen) following manufacturer's instructions. Following PCR amplification (see **Supplementary Table 5**) gel purified DNA fragments were subcloned into TOPO TA vector (Invitrogen) and transformed into TOP-10 bacterial cells. 10 clones per cell line were sent for sequencing in both forward and reverse directions (M13F/R).

**Southern Blotting** 10 μg of restriction enzyme digested genomic DNA was separated on a 0.7% agarose gel, transferred to a nylon membrane (Amersham) and hybridized

with $^{32}P$ random primer (Stratagene) labeled probes for Col1a1 5′ genomic (external) sequences, Col1a1 3′ (internal) located in 3′ arm of Col1a1 targeting vectors (XhoI-PstI fragment of TOPO2.1 5′mCol1a1 & XbaI-PstI 3′mCol1a1 plasmids), Klf4 (full length Klf4 cDNA from pFUW-tetO-Klf4), c-Myc (full length c-Myc cDNA from pFUW-tetO-c-Myc), Oct4 (EcoRI-PstI exon1 fragment of cDNA from pFUW-tetO-Oct4), and Sox2 (full length cDNA from pFUW-tetO-Sox2 plasmid).

**Recombinant HTNCre protein delivery to Col1a1 2lox 4F2A transgenic ES cells:** Transgenic ES cells were passaged one day prior to Cre protein transduction to obtain a monolayer. ES cell medium containing 10 μM HTNCre protein was prepared after appropriate dilution from 200μM stock and incubated with transgenic ES cells for 20 hours. After this time media was changed to ES cell media and cells were incubated for 1 more day until passaging (1:10,000) for single cell colony formation.

**Quantitative RT-PCR:** Total RNA was isolated using Trizol reagent (Invitrogen). Five micrograms of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA Free RNA kit (Zymo Research). One microgram of DNase I–treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately resuspended in 100 μl of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction in an ABI Prism 7000 (Applied Biosystems) with Platinum SYBR green qPCR SuperMix-UDG with ROX (Invitrogen).

Equal loading was achieved by amplifying GAPDH mRNA and all reactions were performed in triplicate. Primers used for amplification are listed in **Supplementary Table**

5. Data were extracted from the linear range of amplification. All graphs of qRT-PCR data shown represent samples of RNA that were DNase treated, reverse transcribed and amplified in parallel to avoid variation inherent in these procedures. Error bars represent s.d. of the mean of triplicate reactions.

**Col 4F2A PCR genotyping**: PCR performed using three sequencing primers (Col1a1 frtA F, frtB R, 4F2A R) listed in Supplementary Table 5. PCR reaction: 95C°/1min (1 cycle), 94 C° /30'', 70 C°/45'' (2 cycles), 94 C°/30'', 68 C°/45''(5 cycles), 94 C°/20'', 66 C°/1min (29 cycles); 4 C°. Wild-type Col1a1 is ~ 300bp, targeted Col1a1 4F2A product is ~ 550bp.

**REFERENCES CITED:**


1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
2. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-317 (2007).
3. Wernig, M. et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-324 (2007).
4. Maherali, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55-70 (2007).
5. Hanna, J. et al. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250-264 (2008).
6. Wernig, M. et al. A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol* **26**, 916-924 (2008).
7. Maherali, N. et al. A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* **3**, 340-345 (2008).
8. Hockemeyer, D. et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3**, 346-353 (2008).
9. Markoulaki, S. et al. Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nat Biotechnol* **27**, 169-171 (2009).
10. Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121**, 465-477 (2005).
11. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23-28 (2006).

12. Carey, B.W. et al. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc Natl Acad Sci U S A* **106**, 157-162 (2009).
13. Soldner, F. et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964-977 (2009).
14. Kaji, K. et al. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* **458**, 771-775 (2009).
15. Woltjen, K. et al. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766-770 (2009).

**ACKNOWLEDGEMENTS**: B.W.C. would like to thank Chris Lengner and Lorenna Lee-Houghton for advice and assistance on somatic cell isolation techniques. We thank Frank Edenhofer for a kind gift of purified recombinant HTN-Cre protein and Jessie Dousman and Ruth Flannery for mouse husbandry. J.H is supported by a fellowship from the Helen Hay Whitney Foundation. R.J. is supported by grants from the NIH: 5-RO1-HDO45022, 5-R37-CA084198, and 5-RO1-CA087869.

**COMPETING INTERESTS**: RJ is an advisor to Stemgent and a co-founder of Fate Therapeutics.

**AUTHOR CONTRIBUTIONS:** B.W.C designed and performed experiments, analyzed data and wrote the paper. S.M. performed all blastocyst injections. J.H. assisted in experiments. C.B. assisted in design of Col1a1 targeting vectors. R.J. participated in experimental design, interpreted data and assisted in writing the paper.

## FIGURE LEGENDS

**FIGURE 1. Reprogramming somatic cells from transgenic Col1a1 4F2A mice.** (**a**) A targeting vector to the 3′ UTR was utilized to deliver four murine reprogramming factors (Oct4, Sox2, Klf4, and c-Myc) as a single dox-inducible polycistronic transgene [11,12].

Red triangles indicate loxP sites. pA indicates poly-adenylation sequence. TetO is tetracycline operator minimal promoter. (**b**) Southern analysis of DNA obtained from Neo-resistant V19 ES cell colonies was digested with XbaI and probed for correct targeting of the Col1a1 3′ UTR using a 5′ external Col1a1 probe to the genomic sequence outside the targeting vector homology arms [11]. Lines #2 and #11 were correctly targeted as indicated by hybridization to 5.3 kb band (untargeted allele ~ 4.7 kb). (**c**) Col1a1 4F2A iPS cells express pluripotency markers. Bright field (BF) images and immunostaining for pluripotency markers alkaline phosphatase (AP), SSEA1, and Nanog. Scale bar, 1mm. (**d**) Oct4 promoter using bisulfite sequencing. Open circles indicate unmethylated and closed circles indicate methylated CpG dinucleotides. Shown are representative sequence analyses from Col1a1 4F2A transgenic MEFs, embryonic stem cells (V6.5), and two Col1a1 4F2A iPSC lines: pro-B-derived #1 and liver-derived line #5B. (**e**) Hematoxylin and eosin staining of teratomas induced after subcutaneous injection of Col1a1 4F2A Liv iPSC #5B and Ker iPSC #2 into SCID mice indicates Col1a1 4F2A iPSCs contribute to all three germ layers. Scale bar, 100μm (**f**) Postnatal chimeric mice detected by agouti coat color from Col1a14F2A MEF iPSC #11 and MAC iPSC #5.

**FIGURE 2: iPSCs derived from Col1a1 3F2Atransgenic MEFs expressing a subset of reprogramming factors.** (**a**) Western blot analysis of 293 cells transiently transfected with pFUW-Ubi 3F2A lentivirus. (**b**) Southern analysis of DNA obtained from Neo-resistant V19 ES cell colonies. DNA was digested with XbaI and probed for correct targeting using a 5′ Col1a1 probe. Three lines (# 2, 4, 18) were correctly targeted as indicated by hybridization to the 5.3kb band (untargeted allele ~ 4.7 kb). (**c**) 3F2A MEFs were infected with a Dox-inducible lentivirus carrying a Sox2 cDNA (pFUW-tetO-Sox2)

and transferred to ES cell medium containing Dox after 48 hours. Typical iPSC colonies appeared around day 12 and were mechanically passaged at day 25. Scale bar, 1mm. (**d**) Immunostaining for AP, SSEA1 and Nanog of iPSC line #34. DAPI staining is shown in the inset. Scale bar, 1mm. (**e**) Southern blot analysis of 3F2A+S iPS cell clones. DNA was isolated from each iPSC line and digested with XbaI. The presence of a single Sox2 pro-viral copy was detected with a Sox2 cDNA probe for hybridization. A similar sized band was detected in lines #36 and 30 suggesting these cells are sibling subclones from the same infected cell. (**f**) 3F2A+S MEF derived iPS cells give rise to post-natal chimeras. High contribution chimera (as detected by agouti coat color) obtained after injection of 3F2A+S iPSC line #34 into blastocysts. This chimera contained iPSCs that contributed to the germline as shown by agouti pups present after mating to BDF1 females.

**FIGURE 3: Cre-excisable Col1a1 4F2A transgenic ES cells.** (**a**) Targeting vector containing loxP sites flanking the 4F2A transgene. A similar targeting scheme as described in Figure 1a was used. (**b**) Targeting of 3F2A to the Col1a1 locus in R26 M2rtTA transgenic ES cells. Southern analysis of DNA obtained from Neo-resistant ES cell colonies digested with XbaI and probed for correct targeting using a 5′ external Col1a1 probe (compare Fig. 1a). Multiple clones were correctly targeted as indicated by hybridization to the 5.3kb band (untargeted allele ~ 4.7kb). Line #2 was chosen for Cre-induced excision. (**c**) Recombinant cell-permeable HTN-Cre transduced into Col1a1 2lox 4F2A ES cells. Cells were grown in ES cell medium containing HTN-Cre protein for 20 hours. 24 hours later the cells were trypsinized, plated at high dilution (1:10,000) and colonies were picked and expanded before being split into ES medium or ES medium + G418. Bright field images depict a subclone that is G418-sensitive. Scale bars, 1mm. (**d**) Southern blot analysis of HTN-Cre-induced excision of Col1a1 2lox 4F2A transgene in ESCs.

DNA obtained from sister clones of Neo-sensitive ES cell colonies was digested with XbaI and probed using a 5′ external Col1a1 probe (compare Fig. 1a). Two subclones (#4 and 9) do not show hybridization to the ~5.3kb band corresponding to the targeted Col1a1 4F2A after exposure to HTN-Cre protein (left). Oct4 and c-Myc cDNA probes do not hybridize to the expected 5.8kb band (also XbaI digested DNA) corresponding to the Col1a1 4F2A Tg (center and right, respectively).

Figure 1

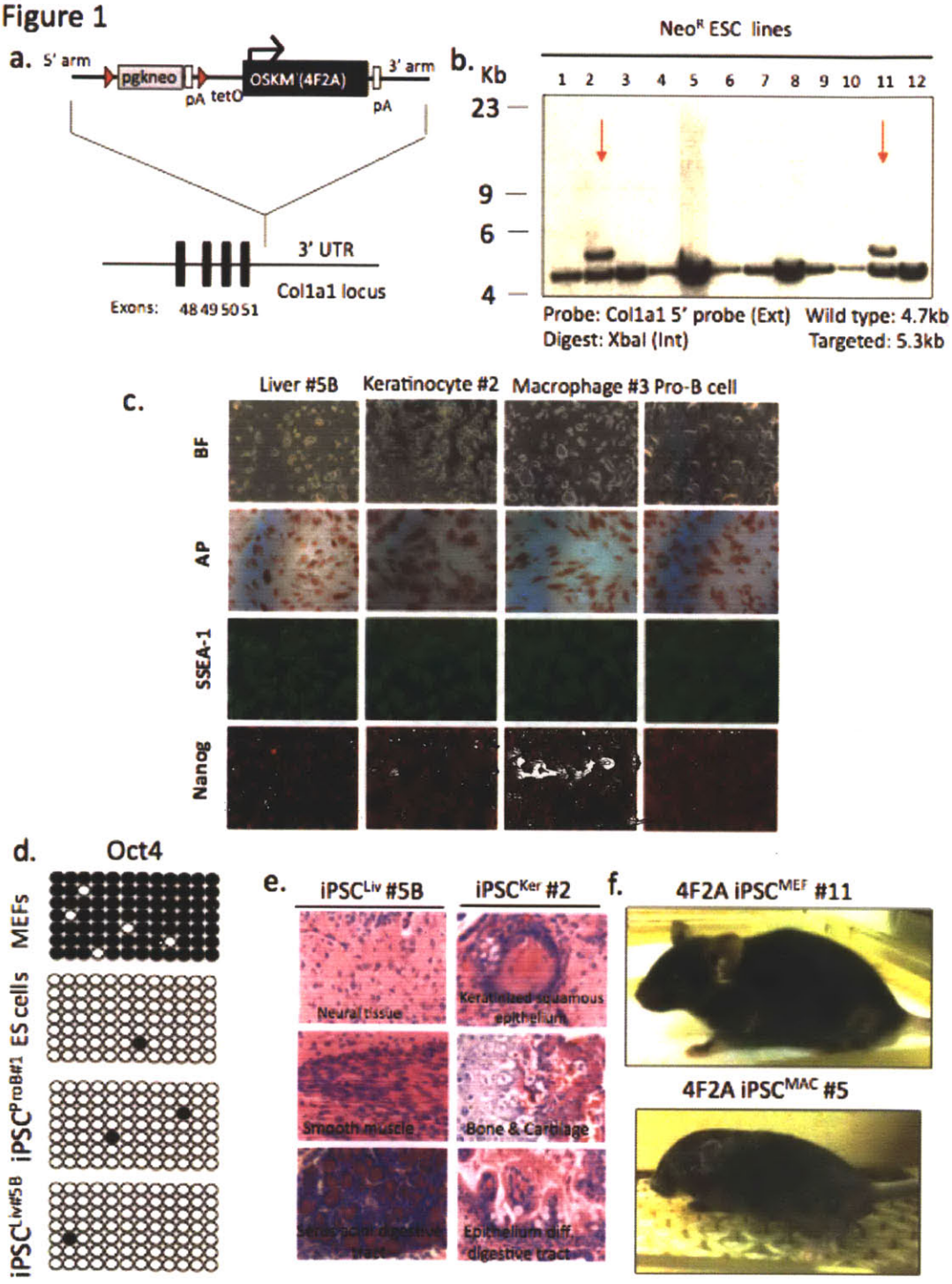

a.
5' arm
pgkneo
pA tetO
OSKM (4F2A)
pA
3' arm
3' UTR
Col1a1 locus
Exons: 48 49 50 51
b.
Neo^R ESC lines
Kb
1 2 3 4 5 6 7 8 9 10 11 12
23
9
6
4
Probe: Col1a1 5' probe (Ext)
Digest: XbaI (Int)
Wild type: 4.7kb
Targeted: 5.3kb
c.
Liver #5B
Keratinocyte #2
Macrophage #3
Pro-B cell
BF
AP
SSEA-1
Nanog
d.
Oct4
MEFs
ES cells
iPSC^ProB#1
iPSC^Liv#5B
e.
iPSC^Liv #5B
iPSC^Ker #2
Neural tissue
Keratinized squamous epithelium
Smooth muscle
Bone & Cartilage
Epithelium diff. digestive tract
f.
4F2A iPSC^MEF #11
4F2A iPSC^MAC #5

Figure 2

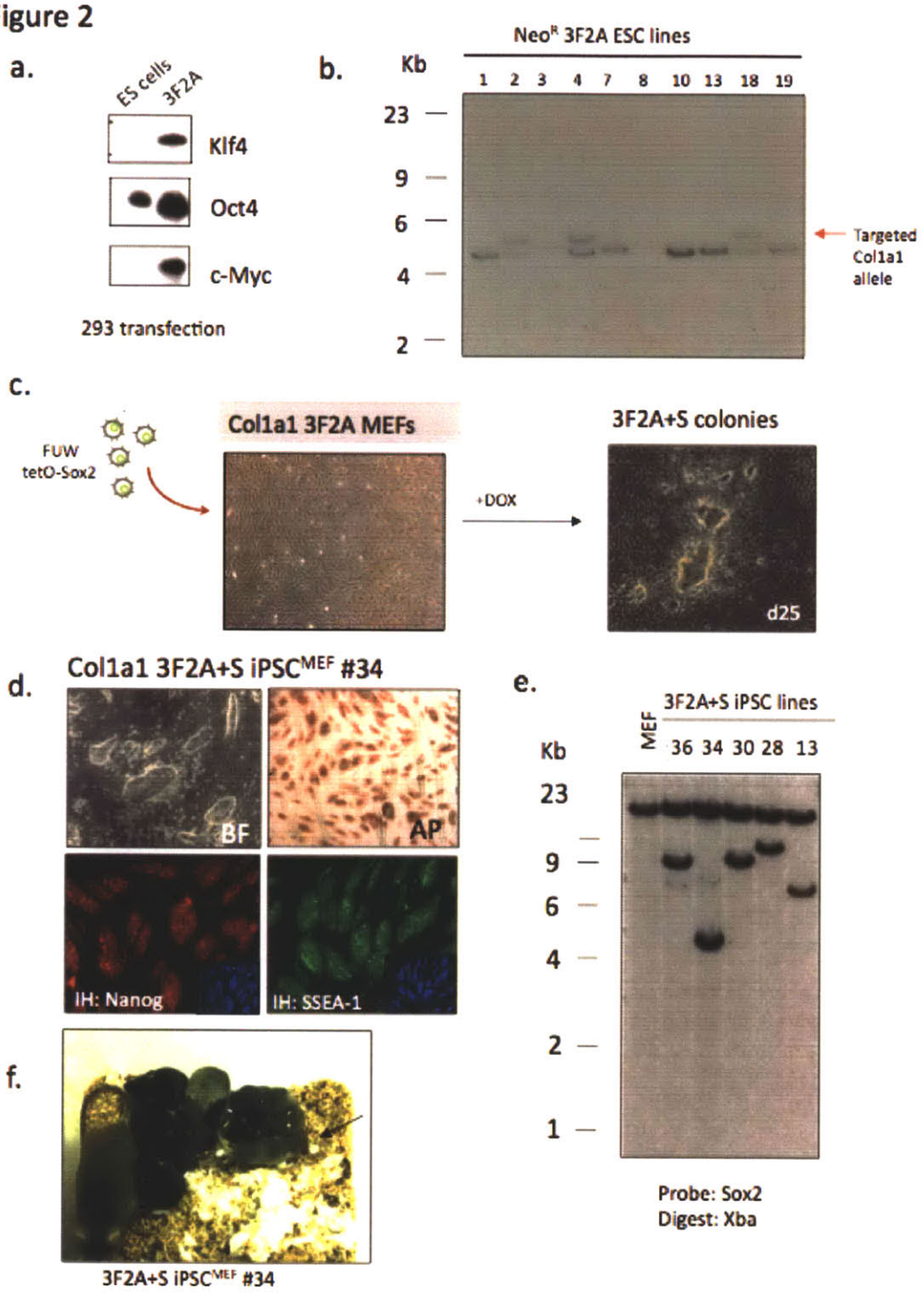
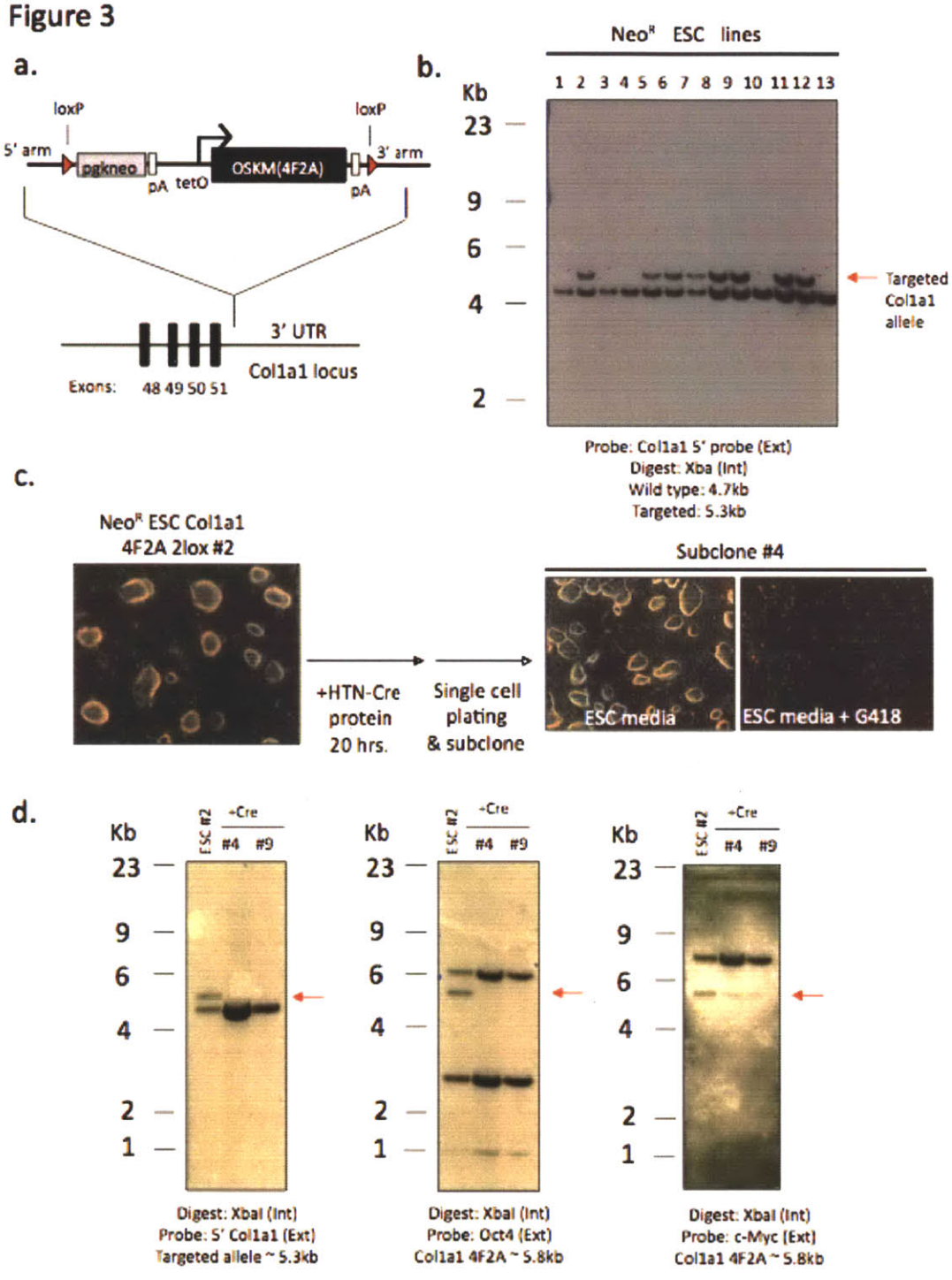
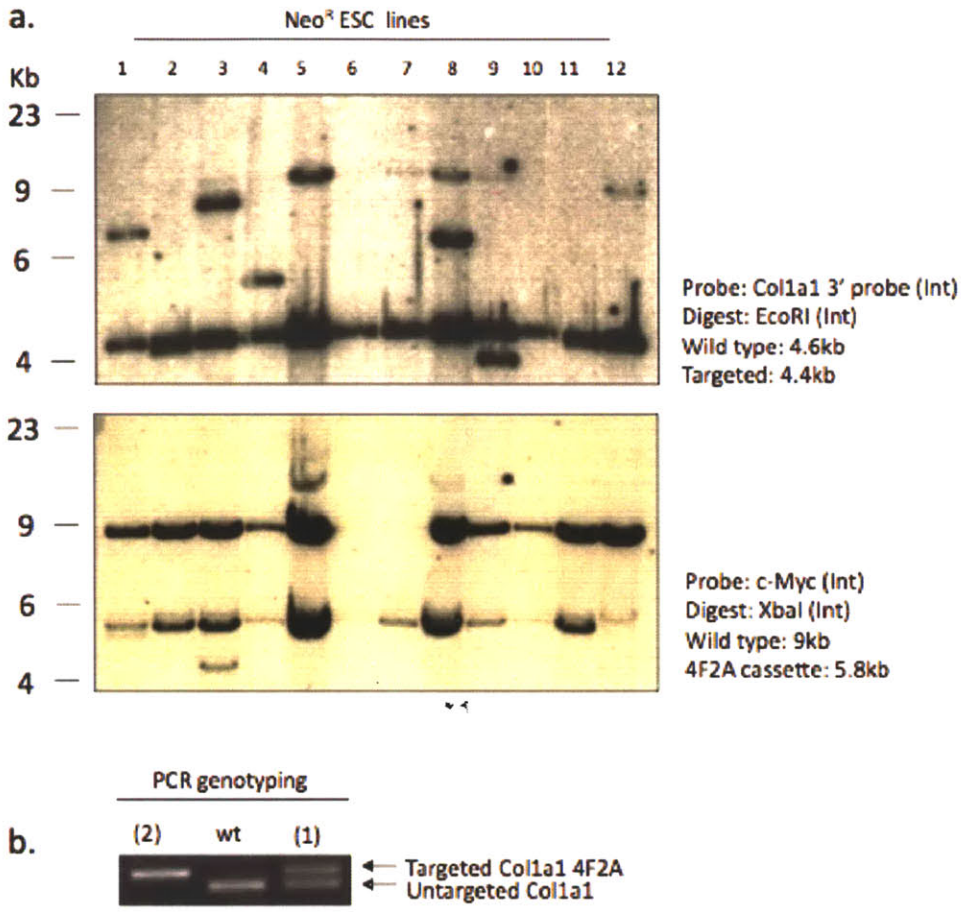

a.
ES cells
3F2A
Klf4
Oct4
c-Myc
293 transfection
b.
Kb
Neo^R 3F2A ESC lines
1 2 3 4 7 8 10 13 18 19
23
9
6
4
2
Targeted Col1a1 allele
c.
FUW tetO-Sox2
Col1a1 3F2A MEFs
+DOX
3F2A+S colonies
d25
d.
Col1a1 3F2A+S iPSC^MEF #34
BF
AP
IH: Nanog
IH: SSEA-1
e.
3F2A+S iPSC lines
MEF
36 34 30 28 13
Kb
23
9
6
4
2
1
Probe: Sox2
Digest: Xba
f.
3F2A+S iPSC^MEF #34

Figure 3

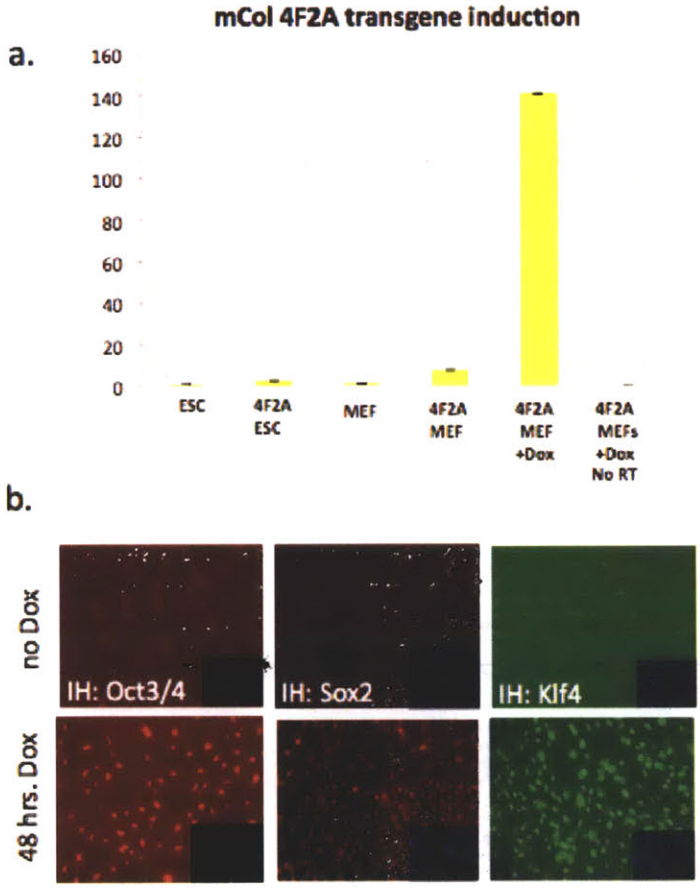

a.
loxP
loxP
5' arm
pgkneo
pA
tetO
OSKM(4F2A)
pA
3' arm
3' UTR
Col1a1 locus
Exons: 48 49 50 51
b.
Neo^R ESC lines
1 2 3 4 5 6 7 8 9 10 11 12 13
Kb
23
9
6
4
2
Targeted Col1a1 allele
Probe: Col1a1 5' probe (Ext)
Digest: Xba (Int)
Wild type: 4.7kb
Targeted: 5.3kb
c.
Neo^R ESC Col1a1 4F2A 2lox #2
+HTN-Cre protein 20 hrs.
Single cell plating & subclone
Subclone #4
ESC media
ESC media + G418
d.
Kb
ESC #2
+Cre
#4
#9
23
9
6
4
2
1
Digest: XbaI (Int)
Probe: 5' Col1a1 (Ext)
Targeted allele ~ 5.3kb
Digest: XbaI (Int)
Probe: Oct4 (Ext)
Col1a1 4F2A ~ 5.8kb
Digest: XbaI (Int)
Probe: c-Myc (Ext)
Col1a1 4F2A ~ 5.8kb

## Supplementary Figure 1



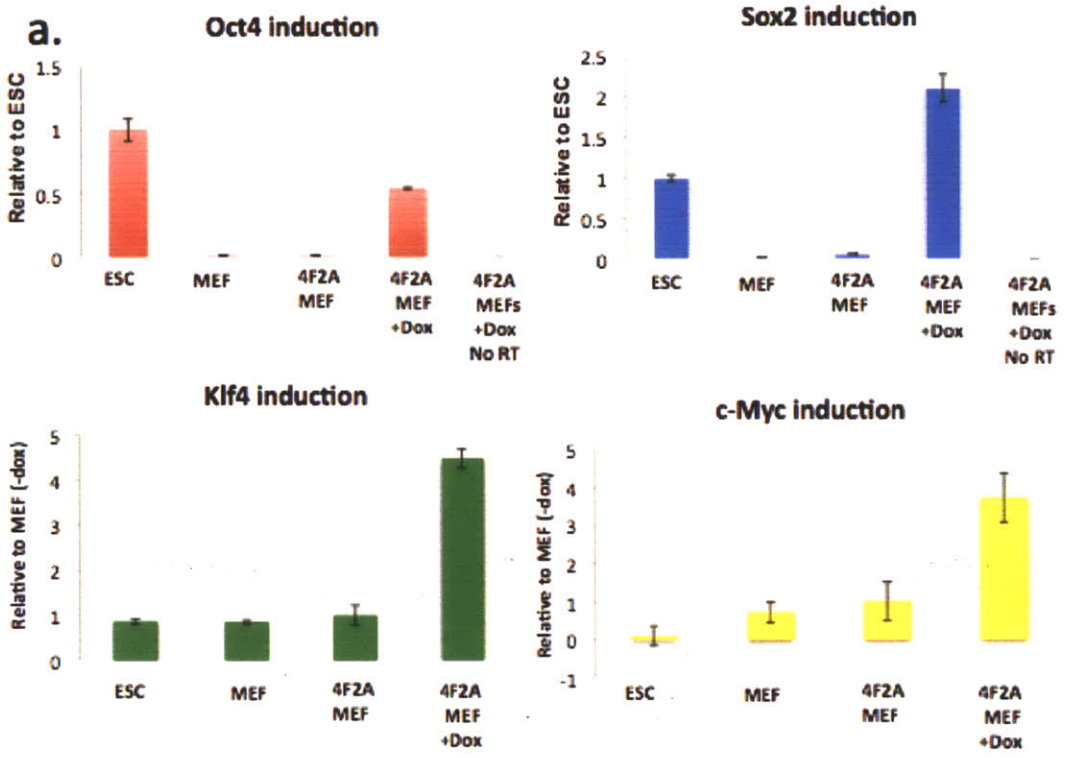

**SUPPLEMENTARY FIGURE 1: Genotyping of Col1a1 4F2A transgenic ES cells and mice.** (a) Southern blot analysis. 10ug of DNA from Neo-resistant ES cells after Col1a1 4F2A targeting was digested with EcoRI. An internal probe (Col1a1 3') was used to detect copy number and correct targeting to the Col1a1 locus. Although the resolution between the targeted and wild-type alleles was not possible, no band other than the genomic was present for clones #2 and #11. These were clones that had the correct band sizes using an external 5' probe for the Col1a1 locus (see Fig. 1a). A second digest (XbaI) and probe using c-Myc confirms a single copy of the 4F2A transgene in clones #2 and #11 (size ~ 5.8kb). (b) PCR genotyping using primers specific for either the targeted or wt Col1a1 locus to detect mice carrying different copy numbers of the Col1a1 4F2A transgene. A targeted allele gave rise to a band size ~ 550bp whereas untargeted allele was ~ 300bp. Primer sequences are listed in Supplementary Table 5.

## Supplementary Figure 2.



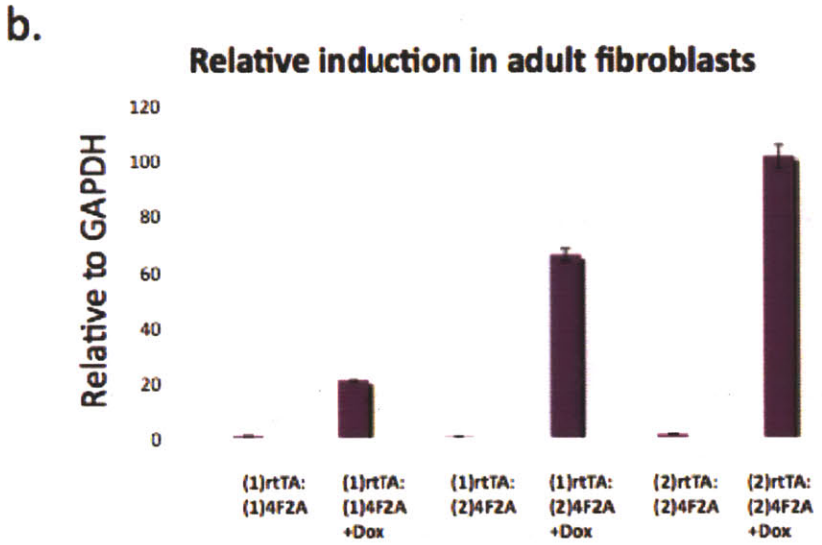

**SUPPLEMENTARY FIGURE 2: Col1a1 dox-inducible expression of 4F2A transgene** (a) qRT-PCR analysis of mRNA induction in single-copy rtTA(1);4F2A(1) transgenic MEFs. Cells were cultured in ES medium + Dox, harvested 48 hours later and RNA was isolated. E2A-cMyc primers were used to detect transgene-specific transcripts. Error bars represent s.d. of the mean of triplicate reactions. (b) Immunostaining for Oct4, Sox2, and Klf4 proteins. Single-copy rtTA(1);4F2A(1) MEFs were cultured in ES media + Dox for 48 hours. Inset pictures are DAPI stain to detect cell nuclei.

## Supplementary Figure 3



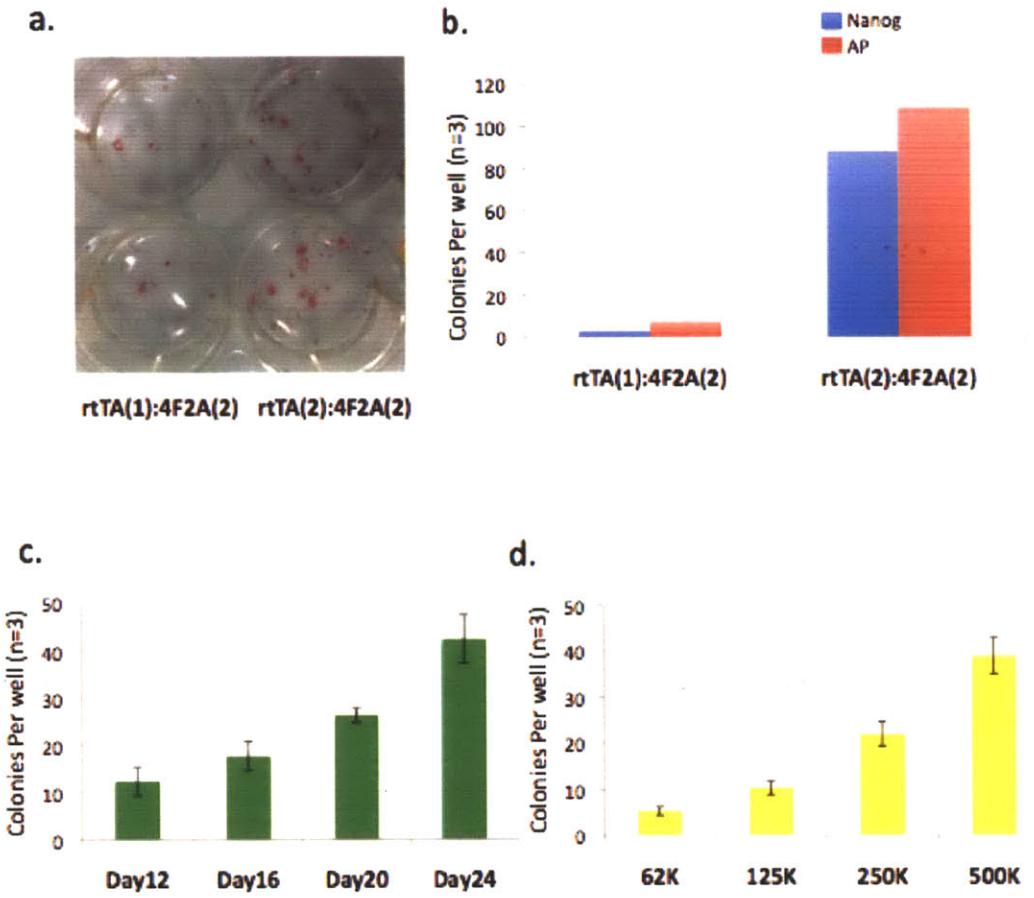

**SUPPLEMENTARY FIGURE 3: Expression of the 4F2A transgene in Col1a1 4F2A (1:1) ratio.** (a) Relative induction of 4F2A Tg using exon-specific primers to determine relative transcript levels compared to ES cells and MEFs. Single copy (1)rtTA;(1)4F2A transgenic MEFs were tested by qRT PCR in the absence or presence of Dox for 48 hours. For both Oct4 and Sox2 the induction was normalized to ES cell levels. For Klf4 and c-Myc the induction was normalized to MEFs. Error bars represent s.d. of the mean of triplicate reactions

## Supplementary Figure 3



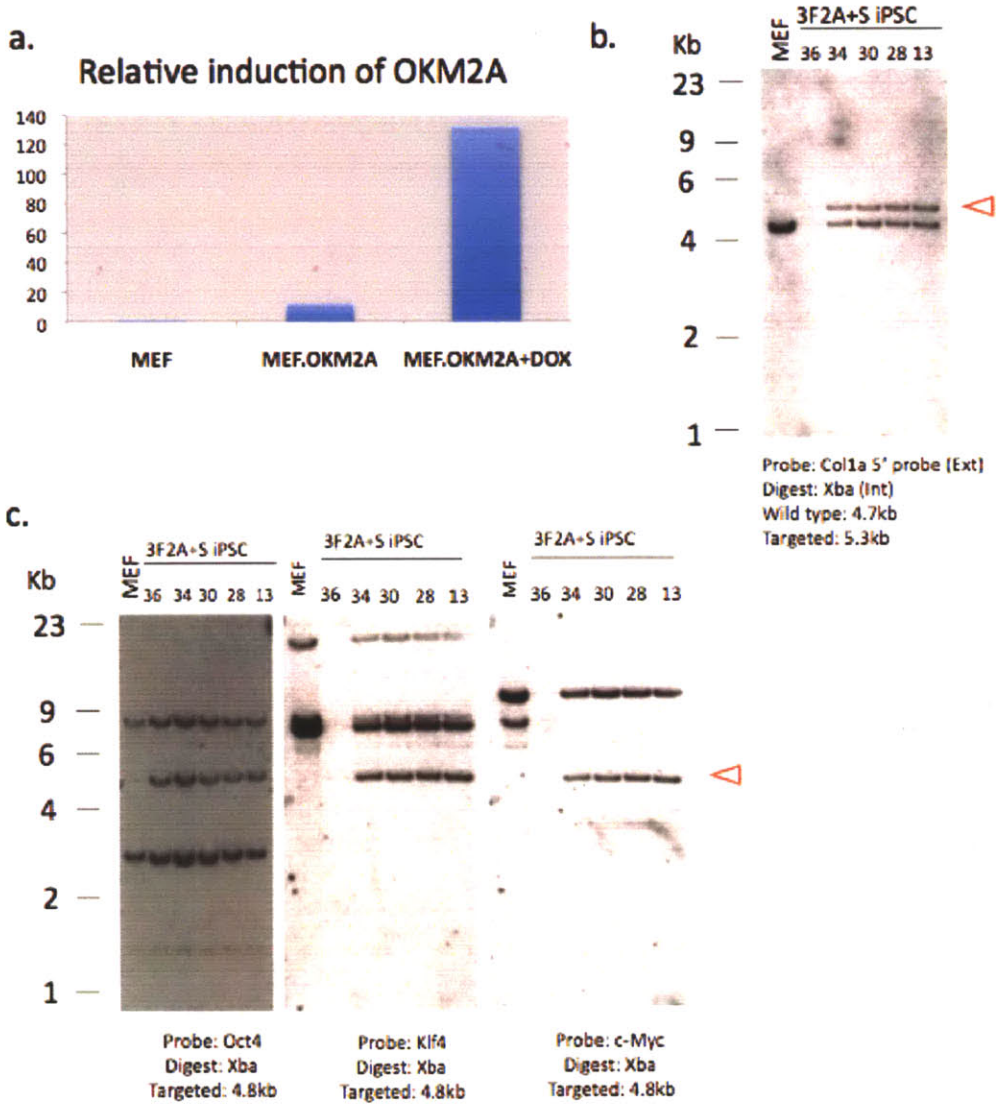

**SUPPLEMENTARY FIGURE 3: Expression of the 4F2A transgene in TTFs derived from mice carrying different numbers of transgene copies.** (b) Effect of transgene copy number on expression. E2A-cMyc primers were used to detect transgene-specific transcripts. Adult tail-tip fibroblasts were isolated from mice carrying 1 copy of each transgene rtTA(1):4F2A(1), 1 copy of rtTA and 2 copies of 4F2A rtTA(1):4F2A(2), or 2 copies of each rtTA(2):4F2A(2) were grown in Dox containing medium for 48 hours and the expression level was quantified by qRT PCR. Error bars represent s.d. of the mean of triplicate reactions.

## Supplementary Figure 5




**SUPPLEMENTARY FIGURE 5: Efficiency of reprogramming Col1a1 4F2A pro-B cells.** (a) Alkaline phosphatase (AP) staining of pro-B-derived colonies on day 25 after transgene induction. Reprogramming was performed in two distinct mouse transgene ratios (1:2 and 2:2). From each strain CD19+ cells were harvested from bone marrow and plated at similar densities ($1x10^6$) per well. On day 20 Dox media was exchanged for ES media without Dox. (b) Quantitative summary of pro-B iPSC colonies as detected by AP (red) and Nanog immunostaining (blue). $1x10^6$ bone marrow derived CD19+ cells were plated into 12-well plates (n= # wells per genotype). Dox containing media was removed on day 20 and three wells were stained and counted respectively on day 25. Error bars represent standard deviation. (c) Time-dependent increase in pro-B iPSC colonies using Col1a1 transgenic mice rtTA(2):4F2A(2). $5x10^5$ bone marrow derived CD19+ cells were plated into 12-well plates (n= # wells per cell density). Wells were stained for Nanog protein and counted on indicated day after 4F2A transgene induction. Error bars represent standard deviation.
(d) Cell density-dependent increase in pro-B iPSC colonies using Col1a1 transgenic mice rtTA(2):4F2A(2). Indicated donor cell numbers were plated into 12-well plates (n = # wells per cell density). Dox containing media was removed on day 20. Wells were stained for Nanog protein and counted on day 25. Error bars represent standard deviation.

## Supplementary Figure 6




**SUPPLEMENTARY FIGURE 6: Dox–induced transgene expression in 3F2A MEFs and characterization of 3F2A+S iPSCs (a)** qRT-PCR analysis of 3F2A MEFs. Primers specific for the OKM-2A Tg (*E2A-Myc* see Methods) were used to detect transgene expression in MEFs cultured in the presence of Dox. Values are normalized to GAPDH and relative to non-transgenic MEFs which were set as "1". **(b)** Southern blot analysis of DNA from 3F2A+S MEF iPSC. DNA was isolated and digested with XbaI to confirm presence of the targeted Col1a1 allele in iPS cell lines. An external 5' probe was used and hybridized to the expected 5.3kb band. **(c)** Southern blot analysis on 3F2A+S MEF iPSC. DNA was isolated and digested with XbaI to confirm the presence of a single OKM-2A transgene. The expected 4.8kb band was detected with probes for Oct4, Klf4, or c-Myc.

Supplementary Table 1. Aging of mice carrying different transgene copies

| genotype | MICE(n) | Age (weeks) |
|---|---|---|
| Col1a1$^{4F2A}$(1); R26$^{rtTA}$ (1) | 9 | 31 |
| Col1a1$^{4F2A}$(1); R26$^{rtTA}$ (2) | 4 | 29 |
| Col1a1$^{4F2A}$(2); R26$^{rtTA}$ (1) | 4 | 25 |
| Col1a1$^{4F2A}$(2); R26$^{rtTA}$ (2) | 2 | 6 |
| Col1a1$^{4F2A}$(1); R26 $^{CAG\text{-}rtTA}$ (1) | 7 | 25 |

**Supplementary Table 1:** Aging of mice carrying different combinations of transgene copies. Animals were carrying different combinations of transgenes were generated as described in Supplementary Fig. 1 and aged. So far no tumors or other health issues have been noted.

**Supplementary Table 2.** iPSC derivation from different donor cell types

| GENOTYPE | MEFs | TTFs | Keratinocyte | MSCs | Liver | Pro-B cells | Macrophage | Intestinal Epith. |
|---|---|---|---|---|---|---|---|---|
| 1. rtTA (1): 4F2A (1) | + | - | - | - | - | - | - | - |
| 2. rtTA (2): 4F2A (1) | ND | - | + | - | + | - | - | - |
| 3. rtTA (1): 4F2A (2) | ND | + | ND | - | + | + | + | - |
| 4. rtTA (2): 4F2A (2) | ND | + | ND | + | + | + | + | + |
| 5. CAG-rtTA (1): 4F2A (1) | ND | - | + | - | ND | - | - | - |

**Supplementary Table 2:** Summary of experiments to derive iPS cells from Col1a1 4F2A mouse strains carrying different numbers of transgene copies. Animals with different combinations of rtTA and 4F2A transgenes were generated by interbreeding of double heterozygous animals (rtTA(1):4F2A(1)). Cells from different tissues were plated in Dox containing medium on 12-well or 6-well plates and independent iPSC lines were established by picking single colonies from separate wells explanted. "+" indicates successful iPSC generation and "-" indicates failure to obtain iPSCs from the respective donor cells. MEFs, mouse embryonic fibroblasts; TTF, tail-tip fibroblasts; MSCs, mesenchymal stem cells. Between 8-24 colonies were picked for each cell type. ND = not done.

**Supplementary Table 3.** Diploid blastocyst injections of transgenic iPSC and ESC lines

| Cell line | Genotype | | Blastocyst Injected | Pups | | |
|---|---|---|---|---|---|---|
| | Col1a1 | M2rtTA | | Chimeric /total | Chimerism (%) | Germline |
| ESC OSKM(4F2A) | +/- | +/- | 40 | 9/13 | 75-100 | 4/5 |
| iPSC$^{MEF}$ 4F2A #11 | +/- | +/- | 45 | 8/16 | 10-60 | ND |
| iPSC$^{MEF}$ OKM-2A+S #34 | +/- | +/- | 26 | 4/5 | 30-90 | 2/4 |
| iPSC$^{VAC}$ 4F2A #M5 | +/+ | +/+ | 36 | 2/5 | 50-100 | 1/2 |

**Supplementary Table 4.** Summary of iPSC pluripotency tests

| Source of cells | ESC markers | Teratoma | Post-natal chimera | Germline |
|---|---|---|---|---|
| Embryonic Fib. (4F2A) | + | + | + | ND |
| Embryonic Fib. (OKM+S) | + | ND | + | + |
| Adult Fib. | + | ND | ND | ND |
| Keratinocyte | + | + | ND | ND |
| Mesenchymal stem cells (MSCs) | + | ND | ND | ND |
| Macrophages (CD11b+) | + | ND | + | + |
| ProB cells (CD19+) | + | ND | ND | ND |
| Liver | + | + | ND | ND |
| Intestinal Epithelium | + | ND | ND | ND |

**ESC markers**: positive for pluripotency markers (AP, SSEA1, Nanog)

**Teratoma**: gives rise to all three germ layers in teratoma assay of pluripotency

**Post-natal chimera**: contributes to postnatal chimera as detected by coat color (agouti)

**Germline**: indicates contribution of iPSC genome to germline

# Chapter 4

## Reprogramming factor expression and stoichiometry influence the epigenetic and biological properties of iPS cells

Bryce W. Carey, Styliani Markoulaki, Jacob Hanna, Kabibi Ganz, Yossi Bouganim, Jongpil Kim, Menno Creyghton, Grant Welstead, Qing Gao, and Rudolf Jaenisch

ABSTRACT:

**We used a genetically defined transgenic system to identify parameters affecting the epigenetic and pluripotent state of iPS cells. We found that high expression of Oct4 and Klf4 combined with lower expression of c-Myc and Sox2 significantly affect the epigenetic and pluripotent state of iPS cells, the majority (80%) of which generated all-iPSC mice by 4n complementation and maintained normal imprinting at the *Dlk1-Dio3* locus. Moreover, LOI at Dlk1-Dio3 did not strictly correlate with reduced pluripotency. Our results argue that idiosyncratic conditions during the reprogramming process profoundly influence the epigenetic state and the pluripotent potential of iPS cells. Thus, it may be problematic to define a "generic" state of iPS cells in efforts comparing the pluripotent potential and epigenetic state of iPS and ES cells.**

**NOTE**: This manuscript is under consideration for publication. We are currently performing additional experiments for resubmission.

**INTRODUCTION:**

Direct reprogramming generates induced pluripotent stem cells (iPSCs) with the molecular profile and developmental potential of embryonic stem cells (ESCs), yet several studies have suggested that subtle differences in gene expression or chromatin modifications yield iPS cell lines with reduced pluripotency when compared to ESCs. Tissue-specific iPSCs have been reported to harbor "epigenetic memory," which correlates with the ability to give rise to somatic cells of their tissue-of-origin more efficiently than other lineages [1-3] arguing the donor cell-of-origin carries a critical influence on the functionality of iPSCs. The identification of molecular markers associated with higher degrees of pluripotency suggested factor-mediated reprogramming is associated with frequent (>95% of isolated lines) aberrant silencing of maternally expressed genes at a single imprinted locus (*Dlk1-Dio3*). The expression state of transcripts from within the locus was reported to be a predictor of the ability to give rise to high contribution chimeras and all iPSC mice in 4n-complementation assays, the most stringent assay for pluripotency [4,5]. These data argue faulty or incomplete reprogramming is a common outcome of factor-mediated reprogramming.

While the successful generation of "all iPSC mice" suggests at least some iPSC lines can pass the most stringent test of pluripotency [6-8], it remains unclear whether technical or biological parameters influence the molecular and biological properties of iPSCs. Comparisons between ESCs and iPSCs often lack proper ESC controls and to date it has been difficult to simultaneously control variables in reprogramming experiments such as complete reactivation of endogenous pluripotency genes, basal reprogramming vector expression [9], genetic background of donor cells, unique reprogramming factor cocktails [10], and distinct proviral integrations [11].

In this study we used a transgenic system that eliminates many if not all of these variables in the generation of iPSCs. Our results demonstrate that even slight variation in transgene stochiometry or reprogramming factor protein levels, leave a major impact on

the epigenetic and biological characteristics of iPS cells and raise the possibility that the epigenetic state and pluripotency of iPSCs is determined by idiosyncratic events during the generation of individual iPS cells.

## RESULTS

### DERIVATION OF TISSUE-SPECIFIC iPSCs:

To generate iPSCs under highly defined conditions from multiple somatic tissues of origin, we isolated cells from a previously published mouse model of factor mediated re-programming [12]. Briefly, these mice carry a single polycistronic vector carrying Oct4, Sox2, Klf4, and c-Myc (herein termed "OSKM-Jae") under the control of a doxycycline-inducible promoter that was inserted into the collagen type 1α locus (Col1a1) in gene-targeted ESCs expressing the reverse tetracycline inducible M2rtTA (rtTA) from the ROSA26 promoter [13]. To test whether basal vector expression has an effect on the reprogrammed cell state iPSCs were derived from the strain carrying an excisable reprogramming factor cassette (2lox OSKM-Jae). Cells derived from four different somatic tissues: liver (Liv), CD19+ proB cells (pB), whole Brain tissue (B) and tail-tip fibroblasts (TTF) were isolated and induced with DOX under optimized conditions for iPSC derivation described previously **(Fig. 1a)**[13]. After 2-3 weeks colonies were mechanically picked from independent wells and DOX-independent iPSC lines established after several passages. All iPSC lines stained positive for pluripotency markers alkaline phosphatase (AP), SSEA-1 and Nanog (**Fig. 1b**). In addition to Oct4 and Nanog, expression of pluripotency markers such as Rex1 (or Zfp42), Esrrb, and Tcf3 were comparable among all nine iPSCs and maintained at levels similar to control ES cell lines as measured by qRT-PCR (**Fig. 1c**). iPSC lines from each somatic lineage generated differentiated tissue of all three germ layers in teratoma assays (**Fig. 1d, Supplemental Fig. 1**) and were able to generate adult chimeras (**Supplemental Table 1**).

Since all nine lines had been derived from the Col1a1 2lox OSKM-Jae mouse excision of the OSKM-Jae cassette was achieved by infection with an adenovirus carrying Cre re-combinase and the green fluorescence protein (GFP), and sorting GFP+ cells 24 hours post-infection then plating at clonal cell concentrations to isolate and analyze independent sub-clones. Southern blot (**Fig. 1e**) and PCR analyses (**Fig. 1f, Supplemental Fig. 2a**) confirmed

precise excision of the vector from the Col1a1 genomic locus. In addition Col1a1 factor-free iPSCs (1lox iPSCs) are distinguishable from factor-containing cell lines as transcript levels were significantly reduced as detected by qRT-PCR (**Supplemental Fig. 2b**). These detectable differences between the factor-containing iPSCs and the factor-free daughter clones are known to have quantifiable changes in global gene expression that may have additional phenotypic consequences during differentiation [9].

**"ALL iPSC" MICE FROM FACTOR-FREE iPS CELLS:**

In agreement with previous studies all iPSCs derived with the Col1a1 OSKM-Jae transgene generated teratomas and adult chimeras after blastocysts injection some of which were shown to contribute to the germline. To determine whether these OSKM-Jae iPSC lines could generate full-term "iPSC mice" we performed tetraploid (4n) complementation assays. Two newly derived ESC controls from a similar breeding scheme to the tested iPSCs were used as a benchmark (Col1a1 OSKM-Jae ES cell lines #1B, #1C). Both ES cell lines (#1B,C) gave rise to live born mice with efficiencies ranging from ~ 1.67 to 3% respectively (**Fig. 2a**) two of which survived to adulthood (#1C). Both were male and had uniform agouti coat color with genotyping confirming the origin of ESC mice (data not shown). To test the potential of tissue-specific iPSCs to generate all iPSC mice we injected all nine independent factor-free iPSCs into 4n-blastocysts and compared their ability to produce live-born mice to ES cell controls. Upon caesarean section 19.5 d.p.c. we obtained breathing pups, termed iPSC mice, with normal morphology in seven out of nine (~ 78%) lines from different somatic origins (**Fig. 2b, Table 1**). Overall efficiencies ranged from 0.5-7.5% (avg. 2.2%), which is similar to ES and SCNT-ES cells as well as other reports with embryonic or adult fibroblast derived iPSC mice [6-8, 14-16]. Most newborns were sacrificed after breathing was established while some gave rise to healthy pups that survived to adulthood (BC2 Ad.5), all of which were male and had a uniformly agouti coat color (**Fig. 2b**).

To exclude minor contributions of tetraploid host cells we utilized a PCR-based approach to detect DNA from host blastocyst-derived cells. Our PCR protocol showed three distinct products depending on whether the Col1a1 locus was untargeted, targeted with OSKM-Jae, or contained a factor-free 1lox allele with a detection limit of ~ 1-5% following a

DNA titration curve (**Fig. 2c**). PCR of DNA from different tissues of six "all iPSC" mice showed only the Col1a1 1lox product with no detectable contribution from host blastocysts cells (**Fig. 2d**). Similar results were observed with tail tissue from BC2 Ad.5 iPSC mice (**Fig. 2d**). These data support the conclusion that OSKM-Jae-induced factor-free iPSCs are competent to generate all-iPSC mice by tetraploid complementation at an efficiency that was similar to that of genetically similar ES cells.

To determine whether the presence of the OSKM-Jae vector interfered with the potential to generate all iPSC mice we injected vector-containing parental iPSCs into 4n-blastocysts. As shown in **Table 1**, five out of the nine lines were able to produce all iPSC mice (avg. 2%) (**Supplemental Fig. 3a**). Most mice appeared normal with a few containing herniated umbilical cords or noticeably larger size (**Supplemental Fig. 3b**). DNA obtained from tissues of sacrificed mice as well as tail tissue from two iPSC adult mice confirmed their origin by PCR with a detection limit of 5% (**Supplemental Fig. 4a**). **Supplemental Fig. 4b** shows that both adult mice contribute the targeted Col1a1 allele to 100% of the genotyped F1 pups (57/57) arguing these are healthy and fertile adult mice.

To examine more closely the general health and physiology of mice generated from pluripotent cells, hematological analyses as well as tests for enzyme secretion indicative of organs such as liver, pancreas, brain and muscle were performed at three months of age. All eight mice that had been derived from germline transmission, ESCs or iPSCs had similar health parameters in 17 categories. (**Supplemental Fig. 5**).

These data indicate that the majority of vector-free and vector containing iPSCs isolated from different tissues were able to generate all iPSC mice after injection into 4n blastocysts which is the most stringent assay available for pluripotency. The efficiency between ES cells and iPSCs was very similar (2.5 and 2.1% respectively, p=0.3071) (**Fig. 2c**). In our experiments the OSKM-Jae transgene did not appear to greatly influence to ability to generate all iPSC mice (2.0 vs. 2.2%). In addition, tissue-of-origin influences on the pluripotency of iPSCs appear to be negligible in 4n-complementation assays using the Col1a1 OSKM-Jae reprogramming system.

## SILENCING OF DLK1-DIO3 LOCUS IS NOT A STRICT MARKER OF REDUCED PLURIPOTENCY

Although residual expression of the four factors did not greatly impair the ability to generate all iPSC mice we observed that some lines failed in this assay. Two recent reports found that aberrant epigenetic silencing of a single imprinted locus, a region on chromosome 12qF1 flanked at each end by the *Dlk1-Dio3* genes respectively, correlated with the inability of iPS cell lines to generate all iPSC mice by 4n-complementation [4,5]. To determine whether this could faithfully predict the failure of iPSCs from our system to produce all iPSC mice, we analyzed expression of transcripts from within the *Dlk1-Dio3* region. We tested two maternally expressed imprinted RNAs, the noncoding RNA *Gtl2* (also known as *Meg3*) and the small nucleolar RNA *Rian*. Both genes were expressed at levels similar to ES cell control in a majority of lines (6/9), while a minority (3/9) exhibited reduced expression (**Fig. 3a**). Several lines that had repressed the *Dlk1-Dio3* locus exhibited high contribution chimeras with some contributing to the germline (iPSC clone pBL4) (**Fig. 3b**).

To further analyze the molecular properties of the *Dlk1-Dio3* locus in Col1a1 OSKM-Jae iPSCs we examined the extent of DNA methylation at a key regulatory element, the intergenic germline-derived differentially methylated region (IG-DMR), which is a control element for the imprinted region [17,18]. Using methylation-sensitive Southern blotting we tested two ES cell clones together with six iPSC lines that either showed reduced or ESC-like levels of both *Gtl2* and *Rian*. Pluripotent cell lines with ESC-like expression of *Gtl2* and *Rian* showed both methylated and unmethylated IG-DMR alleles (**Fig. 3c**) in two independent methylation-sensitive digests (*HpaII* or *HhaI*). However, only methylated IG-DMR alleles were detected in two iPSC lines (Livp2E2, pBL4) with reduced levels of *Gtl2* and *Rian* expression, while the third (BE1) did not show a gain in methylation at the CpGs tested by Southern. With the exception of iPSC line BE1, we observed that the majority of lines with repressed transcripts within *Dlk1-Dio3* locus showed a gain of DNA methylation at CpGs within the IG-DMR locus, whereas iPSCs with ESC-like expression did not. Similar analyses on additional iPS cell lines derived from keratinocytes and MEFs further demonstrated that the majority of iPSCs derived using the OSKM-Jae transgene preserved normal imprinting of the *Dlk1-Dio3* locus (**Supplemental Fig. 6**).

Two previous papers reported that silencing of the *Dlk1-Dio3* imprinted region correlated strictly with reduced chimerism, efficiency of contribution to the germline, and an inability to generate late stage embryos or all iPSC mice in 4-complementation assays [4,5].

To further assess the developmental potential of the Col1a1 OSKM-Jae iPSCs we performed 2n blastocyst injections to generate adult chimeras from all nine lines to analyze chimerism based on agouti coat color (n=44) dividing the chimeras into two groups based on Gtl2 expression (Gtl2-ON or Gtl2-LOW). All lines, independent of the status of *Dlk1-Dio3* had variable contributions with no significant difference between the two groups (**Fig. 3d,e**).

Our results differ from the previous study that demonstrated the predictive value of the imprinting state of the Dlk1-Dio3 locus in determining the potential of iPSC to produce all iPSC mice (**Fig. 3f**). Factor-free iPSC line pBL4 Ad.1 gave rise to all iPSC mice (**Table 1**) and late stage embryos with southern blot analysis of the IG-DMR locus confirming that these mice carried only methylated alleles (**Fig. 3g, Supplemental Fig. 7a,b**). We conclude that reduced expression of genes within the *Dlk3-Dio3* locus does not appear to *strictly* correlate with reduced pluripotency in iPSCs. Our data do not, however, exclude the possibility that iPSCs with loss of imprinting at *Dlk1-Dio3*, can only generate iPSCs after the vectors have been excised.

## STOCHIOMETRY DURING REPROGRAMMING INFLUENCES THE EPIGENETIC STATE OF DLK1-DIO3 LOCUS

It is puzzling that two transgenic mouse strains carrying almost identical transgenes (**Supplemental Fig. 8**) inserted into the same expression locus generated normal iPSCs at significantly different frequency and did not show a strict correlation with Dlk1-Dio3 locus as expected. To investigate the basis for these differences we analyzed the contribution of transgene expression and reprogramming factor stoichiometry in both strains (referred to as "OSKM-Jae" [13] vs. "OKSM-KH" [16]).

In order to rule out the possibility that genetic background or differences in transgene ratios within our system modifies *Dlk1-Dio3* imprinting, we analyzed established blood and liver-derived iPSCs from four genetically distinct mice carrying unique transgene ratios (e.g. heterozygous or homozygous) at the ROSA26 and Col1a1 genes. Indeed, neither had an impact on our results since, similar to our earlier observations, we found six (66%) iPSCs showing ESC-like expression of *Gtl2*, *Mirg*, and *Rian*, and only three (33%) exhibiting reduced expression of these genes (**Supplemental Fig. 9**).

To determine whether measurable differences distinguished these two systems, we tested transgene induction as well as reprogramming factor protein levels from transgenic mouse embryonic fibroblasts (MEFs) derived from the same Col1a1 targeted parental ES cell lines (OSKM-Jae or OKSM-KH). Transgenic MEFs were cultured in DOX and transgene induction levels analyzed by qRT-PCR using primers for Oct4 and Sox2 cDNA. Neither system showed significant differences in transgene RNA levels prior to or upon induction (**Fig. 4a**). To compare the translation efficiency of the two transgenes, we analyzed the protein levels of the reprogramming factors. Surprisingly, western blot analysis showed that reprogramming factor protein levels were significantly different in the two strains with Col1a1 OSKM-Jae induced MEFs exhibiting higher levels of Oct4 and Klf4 and lower levels of Sox2 and c-Myc than those of similarly induced Col1a1 OKSM-KH MEFs (**Fig. 4 b**). Quantification of band intensities after normalization to the loading control, indicates that OSKM-Jae MEFs had ~ 5- and 15-fold higher protein levels of Oct4 and Klf4 and 0.56- and 0.51-fold lower protein levels of Sox2 and c-Myc as compared to OKSM-KH MEFs (**Fig. 4c**). High levels of unprocessed large molecular weight precursor proteins were consistent with inefficient processing to mature functional Oct4 and Klf4 proteins in OKSM-KH cells. (see * in **Fig. 4 b**). These results demonstrate that subtle differences in the vector construction have an important impact on the resulting protein levels.

To examine whether the differences in reprogramming factor protein levels can directly impact the epigenetic state of the iPSCs we analyzed IG-DMR methylation in iPSCs created by induction of OSKM-Jae alone or by additional ectopic expression of Oct4 or Klf4 or of both factors using dox-inducible lentiviruses. For this we derived secondary MEFs from a single Gtl2-ON iPSC line, BC2, by injection into 2n-blastocysts and initiated reprogramming to generate stable DOX-independent iPSCs in all four conditions (OSKM, OSKM+O, OSKM+K, OSKM+O+K). At day 20, twenty-four colonies were picked from independent wells in all conditions and the establishment of DOX-independent iPSCs occurred at varying rates. Induction of OSKM-Jae with addition of Oct4, or both Oct4 and Klf4, increased the rate of DOX-independent iPSCs dramatically, with >80% of colonies picked generating iPSCs independent of reprogramming factors by two passages. This was in contrast to OSKM-Jae alone or OSKM+K, which showed 30% and 50% of picked colonies giving rise to DOX-independent iPSCs at passage 3 (**Supplemental Fig. 10**). Upon isolation of ad-

ditional DOX-independent iPSCs from all conditions, DNA was harvested and analyzed by Southern blot to determine the extent of aberrant DNA methylation at the IG-DMR within the *Dlk1-Dio3* locus. Aberrant methylation occurred at a rate of 44% (OSKM), 42% (OSKM+O), or 42% (OSKM+K) in iPSCs from MEFs. However, additional ectopic expression of Oct4 and Klf4 significantly reduced aberrant silencing to ~ 15% of recovered lines (OSKM+O+K) (p=0.0441 Fisher's exact, two-tailed) (**Fig. 4e,f**) demonstrating that specific reprogramming factor levels during reprogramming can influence the resulting epigenetic makeup at the IG-DMR locus in iPSCs. The specificity of the process is exemplified by the accelerated appearance of stable iPSCs by Oct4 overexpression alone, but was insufficient to reduce aberrant DNA methylation at the IG-DMR locus.

**iPSC MICE DO NOT SHOW ELEVATED MORTALITY FROM TUMORS**

The generation of a large number of transgene-free iPSC lines from adult tissues afforded us the ability to analyze the susceptibility of somatic cells derived from iPSCs to generate tumors in vivo. We additionally chose to analyze long-term health outcomes of mice from the Col1a1 OSKM-Jae strain because, in contrast to other systems, it did not cause increased mortality rates over a 16-month time period even in the presence of transgenes. High contribution chimeric mice (>90%) derived from Col1a1 OSKM-Jae transgenic ESCs survived for up to 16 months without any detectable weakness or presence of tumors in contrast to chimeras generated from Col1a1 OKSM-KH transgenic ESCs confirming previous results [16] (**Supplemental Fig. 11**).

To this end we analyzed long-term survival trends for iPSC chimeras from two mouse cohorts, one for mice generated from factor-containing iPSCs (n=38) and the other for mice derived from factor-free iPSCs (n=36) to observe any elevated rates of tumorigenesis. **Fig. 5a** shows that for up to 16 months iPSC chimeras survived without any detectable tumors or weakness causing death with similar results obtained from factor-free iPSC-derived mice kept so far for up to four months. The origins of iPSC-derived chimeras were from reprogrammed MEFs, CD11b+ macrophages, CD19+ proB cells, adult fibroblasts, whole brain and liver tissue. In mice that were sacrificed no tumors were detected upon post-mortem pathological analysis, and mortality was typically the result of problems found in non-transgenic strains such as extreme dermatitis and in one case, a hernia. **Fig.**

**5b** indicates that cumulative mortality rates were no different in aged mice of the Col1a1 OSKM-Jae mouse colony.

Using the Col1a1-OSKM-Jae reprogramming system we observed normal lifespans in iPSC-derived mice for up to 16 months with ~ 90-100% of mice surviving 12 months. These data suggest that at least during the time of observation of this study, the Col1a1-OSKM-Jae system does not produce adult mice with increased predispositions towards tumorigenesis following somatic cell differentiation.

**DISCUSSION:**

In this paper we used a highly defined inducible transgenic system that eliminates all genetic variables for the generation of iPSCs. We derived multiple iPS cells from several tissues of an adult mouse and demonstrate that the majority are capable of generating high contribution iPSC chimeras as well as "all iPSC mice" in 4n-complementation assays at frequencies similar to control ESCs. Our results argue that tissue-of-origin has a limited if any effect on the ability of iPSCs to differentiate as tested by this most stringent assay for pluripotency. Moreover, only a minority of our iPSC lines silenced transcripts within the imprinted *Dlk1-Dio3* locus and contrary to expectations, this did not preclude the generation of all-iPSC mice. We further characterized differences in factor stoichiometry between reprogramming systems and demonstrate that this can directly influence aberrant DNA methylation within the *Dlk1-Dio3* locus during derivation of iPSCs. Finally, we failed to detect any tumors in mice derived from these iPS cells over the course of 12-16 months.

Our data emphasize the critical importance of factor level and/or stoichiometry as a key determinant of the pluripotent state of iPSCs. The recovery of a high proportion of iPSCs with identical pluripotency as ESCs correlated with higher levels of pluripotency factors such as Oct4 (5-fold) and Klf4 (15-fold) and lower levels of Sox2 (0.5-fold) and c-Myc (0.5-fold) when compared to a system that yielded iPSCs of which only a small fraction were competent to generate all-iPS mice by 4n-complementation [5, 16]. Though the two transgenic systems (OSKM-Jae vs. OKSM-KH) were identical, i.e. polycistronic vector insertion into identical positions in the 3' UTR of the Col1a1 gene under the DOX inducible regulation of the same transactivator carried in the ROSA 26 locus, the biological properties of the iPCSs such as imprinting of the *Dlk1-Dio3* locus, 4n competence and tumor incidence

were very different. Though both, DOX induced and basal transcription levels of the vectors were very similar in both strains, major differences in the absolute and relative reprogramming factor protein levels were observed. The different efficiency of translation and processing of the factors may be explained by subtle differences in vector design: (i) the arrangement of the second and third factor differed (SK vs. KS), (ii) two different 2A sequences were used (P2A vs. F2A) and (iii) the two COOH terminal factors Sox2 and c-Myc were preceded by an IRES sequence in the OKSM-KH mice as opposed to a T2A and E2A sequence preceding Klf4 and c-Myc in the OSKM-Jae mice. Indeed, the high levels of unprocessed large molecular weight precursor proteins in the OKSM-KH strain were consistent with inefficient processing to mature functional Oct4 and Klf4 proteins, potentially due to the F2A peptide used to mediate translation of these proteins.

The elevated rates of LOI at the *Dlk1-Dio3* locus previously reported, suggested that factor-mediated reprogramming was ineffective at maintaining normal imprinting at this locus during reprogramming. We demonstrate that these results are highly system dependent and may be due to the specific expression and/or stoichiometry of particular reprogramming factors. Consistent with this hypothesis we found that enhanced expression of Oct4 and Klf4 in OSKM-Jae MEFs further reduced aberrant methylation at the *Dlk1-Dio3* locus during iPSC isolation.

Of critical importance to the iPS field is the question of how similar or different iPS cells are as compared to ES cells. A number of recent studies have demonstrated epigenetic, genetic and biological differences between iPS and ES cells though, at least in some cases, the functional consequences of these differences have not been clarified [1, 2, 19-22]. However, a major complication in these comparisons is whether and how we can define a "generic" state of ES and iPS cells. The results presented in this study argue that subtle and temporary differences in the level and stochiometry of the reprogramming factors during early stages of iPSC formation can have profound and irreversible consequences for the epigenetic state and the pluripotent potential of iPS cells. This is consistent with the notion that incomplete or imperfect reprogramming is not a fundamental biological problem of factor induced iPSC generation but rather due to technical and often not well-controlled variables. Thus, comparisons between ES and iPS cells may often be influenced by idiosyncratic and

difficult to control differences that occur during the generation of individual pluripotent cells qualifying conclusions about generic differences between ES and iPS cells.

**MATERIALS AND METHODS:**

**Somatic cell isolation and culture:** Somatic organs were isolated from a 6-week-old agouti Col1a1 2lox OSKM-Jae transgenic male mouse heterozygous for the Rosa26 M2rtTA and homozygous at the Col1a1 locus harboring the 2lox OSKM-Jae cassette. Adult tissues were isolated with whole brain tissue being plated immediately into ES cell medium with doxycycline (2 μg/ml). For blood reprogramming whole marrow was isolated from the femur and tibia after removal of the condyles at the growth plate by flushing with a syringe and 30-guague needle containing DMEM+5% FBS (Hyclone, Thermo Fisher Scientific). CD19+ pro-B cells were isolated by MACS cell separation (Miltenyibiotec Cat# 130-052-201) following manufactures instructions. Purified B cell subsets were resuspended in IMDM with 15% FCS as well as IL-4, IL-7, SCF (10 ng/ml each, Peprotech), doxycyline (2 μg/ml) and plated on OP9 bone marrow stromal cells (ATCC). Three days later the medium was changed to ESC medium plus Dox. For isolation of liver cells mice were first perfused with 50 ml HBSS buffer (w/o Ca2+ and Mg2+) then 50 ml HBSS (w/o Ca2+ and Mg2+) containing collagenase (type IV) (Sigma Cat# C5138) (100U/ml). Liver was dissected away from surrounding tissues and dissociated in 10ml DAG media (phenol-red free EMEM Gibco-11054-020 and Bovine serum albumin (BSA) 1g/0.5L) and filtered two times through a sterile 100 μM cell strainer. Liver cell preparations were centrifuged at 30 g for 3 minutes at 4 °C and the cells were washed two times with DAG media and then plated on γ-irradiated MEFs in ES media + Dox.

**Excision of Col1a1 2lox OSKM-Jae cassette in iPSCs**

For Cre-recombinase-mediated vector excision iPSCs were plated into a single well of 12-well plate at high confluency and infected with adenovirus carrying GFP and Cre-recombinase purchased from University of Iowa gene core (http://www.uiowa.edu/~gene/) for 24 hours. Cells were harvested with 0.05% trypsin/ethylenediaminetetraacetic acid (EDTA) solution (Invitrogen), and subjected to FACS sorting (FACS-Arai; BD Biosciences) of

a single cell suspension for EGFP-expressing cells 24 hours after infection. Cells were re-plated at a low density containing ES cell medium. Individual colonies were picked 10 to 14 days after electroporation.

**Col1a1 2lox OSKM-Jae ES cell isolation**

Following protocols described previously [23] we isolated ES cells from blastocyst embryos using ESC derivation media: KOSR (knock-out serum replacement, Gibco, Cat# 1028-028), LIF ESGRO (1× 107 U ESGRO/ml; Chemicon, Cat# ESG1106), Mek1 inhibitor (PD98059; Cell Signaling Technology, Cat# 9900), non-essential amino acids, glutamine solution, pen/strep solution. Embryos were obtained after mating C57Bl6/DBA/129 mixed non-sibling male and female mice heterozygous at ROSA26 and Col1a1 locus (rtTA +/-; OSKM-Jae +/-). For ESC derivation isolated 3.5 d.p.c. blastocysts were plated in 4-well plates containing MEF feeders. Immediately prior to plating, the zona surrounding the blastocyst was removed by brief exposure to acid Tyrode's solution (Sigma, Cat# T-1788). After zona removal, a single blastocyst was placed into a well of a 4-well plate, placed in a $CO_2$ incubator and allowed to attach to the feeder layer. Attachment is observed within 24–48 h. After passaging genotyping was performed on Col1a1 OSKM-Jae ESC lines: #1B is homozygous at the ROSA26 locus (rtTA +/+) and wildtype at Col1a1 (-/-). #1C is homozygous for both transgenes (ROSA26 & Col1a1).

**Col OSKM-Jae PCR genotyping**: PCR performed using three sequencing primers (Col1a1 frtA F, frtB R, 4F2A R) published previously[13]. PCR reaction: 95C°/1min (1 cycle), 94 C° /30", 70 C°/45" (2 cycles), 94 C°/30", 68 C°/45"(5 cycles), 94 C°/20", 66 C°/1min (29 cycles); 4 C°. Wild-type Col1a1 is ~ 300bp, targeted Col1a1 4F2A product is ~ 550bp.

**Quantitative RT-PCR**

Total RNA was isolated using Trizol reagent (Invitrogen). Five micrograms of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA Free RNA kit (Zymo Research). One microgram of DNase I–treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately resuspended in 100 ul of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction in an ABI Prism 7000 (Applied Biosystems) with Platinum SYBR green qPCR SuperMix-UDG with ROX (Invitrogen). Equal loading was achieved by amplify-

ing GAPDH mRNA and all reactions were performed in triplicate. Primers used for amplification were as follows:

*Oct4* F, 5'-ACATCGCCAATCAGCTTGG-3' and

R, 5'-AGAACCATACTCGAACCACATCC-3'

*Sox2* F, 5'-ACAGATGCAACCGATGCACC-3' and

R, 5'- TGGAGTTGTACTGCAGGGCG-3'

*OSKM-Jae (E2A-cMyc)* F, 5'-GGCTGGAGATGTTGAGAGCAA-3' and

R, 5'-AAAGGAAATCCAGTGGCGC

*GAPDH* F, 5'-TTCACCACCATGGAGAAGGC-3' and

R, 5'-CCCTTTTGGCTCCACCCT-3'

*HPRT* F, 5'-GCAGTACAGCCCCAAAATGG-3'

R, 5'-GGTCCTTTTCACCAGCAAGCT-3'

**Southern blotting**

For analysis of OSKM-Jae excision in iPSC lines 10 μg of *XbaI* digested genomic DNA was separated on a 0.7% agarose gel, transferred to a nylon membrane (Amersham). The membrane was pre-hybridized with hybridization buffer (1% bovine serum albumin in 7% SDS and 0.5M sodium phosphate, pH 7.5) at 65°C and hybridized with $^{32}P$ random primer (Stratagene) labeled probes for 5' Col1a1 probe (external) [13] and mouse Klf4 (full length Klf4 cDNA). For analysis of IG-DMR, DNA was digested with *StuI* plus methylation-insensitive *MspI* or methylation-sensitive *HpaII* or *HhaI*. The blot was hybridized with probe M4 (IG-DMR) as previously described [24].

**Western Blotting**

Cell pellets were lysed on ice in Laemmli buffer (62.5 mM Tris-HCl pH 6.8, 2% sodium dodecyl sulphate, 5% β-mercaptoethanol, 10% glycerol and 0.01% bromophenol blue) for 30 min in presence of protease inhibitors (Roche Diagnostics), boiled for 5–7 min at 100°C and subjected to western blot analysis. Primary antibodies: rabbit anti-Oct4 (1:1000, scH-134, Santa Cruz), rabbit anti-Sox2 (1:1000, #2748, Cell Signaling), goat anti-Klf4 (1:1000, AF3158, R&D Biosystems), rabbit anti-c-Myc (1:1000, #9402, Cell Signaling), and rabbit anti-actin (1:2000, A2066, Sigma). Blots were probed with anti-mouse, anti-goat, or anti-rabbit IgG-HRP secondary antibody (1:10,000) and visualised using ECL detection kit (GE

Healthcare).

**Immunofluorescent staining**

Cells were fixed in 4% paraformaldehyde for 20 minutes at 25 °C, washed 3 times with PBS and blocked for 15 min with 5% FBS in PBS containing 0.1% Triton-X. After incubation with primary antibodies against Oct4 (Santa Cruz h-134), Sox2 (R&D Biosystems), mNANOG (Bethyl A300-398A), SSEA1 (monoclonal mouse, Developmental Studies Hybridoma Bank) for 1 h in 1% FBS in PBS containing 0.1% Triton-X, cells were washed 3 times with PBS and incubated with fluorophore-labeled appropriate secondary antibodies purchased from Jackson Immunoresearch. Specimens were analyzed on an Olympus Fluorescence microscope and images were acquired with a Zeiss Axiocam camera.

**Mouse Chimera and Teratoma Formation**

All 4n-injections were performed using B6D2F2 embryos as the iPSC lines and ESCs. Col1a1 2lox OSKM-Jae iPSCs were derived from an agouti mouse and could be identified by coat-color as adults. Diploid or tetraploid blastocysts (94–98 h after hCG injection) were placed in a drop of Hepes-CZB medium under mineral oil. A flat tip microinjection pipette with an internal diameter of 16 μm was used for iPS cell injections. Each blastocyst received 8-10 iPS cells. After injection, blastocysts were cultured in potassium simplex optimization medium (KSOM) and placed at 37 °C until transferred to recipient females. About 10 injected blastocysts were transferred to each uterine horn of 2.5-day-postcoitum pseudo-pregnant B6D2F1 female. Pups were recovered at day 19.5 and fostered to lactating B6D2F1 mothers when necessary. Teratoma formation was performed by depositing 2x10^6 cells under the flanks of recipient SCID or Rag2-/- mice. Tumors were isolated 3-6 weeks later for histological analysis.

**REFERENCES CITED:**


1. Kim, K. et al. Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285-290.
2. Polo, J.M. et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28**, 848-855.
3. Hu, Q., Friedrich, A.M., Johnson, L.V. & Clegg, D.O. Memory in induced pluripotent stem cells: reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation. *Stem Cells* **28**, 1981-1991.

4. Liu, L. et al. Activation of the imprinted Dlk1-Dio3 region correlates with pluripotency levels of mouse stem cells. *J Biol Chem* **285**, 19483-19490.
5. Stadtfeld, M. et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175-181.
6. Zhao, X.Y. et al. iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86-90 (2009).
7. Boland, M.J. et al. Adult mice generated from induced pluripotent stem cells. *Nature* **461**, 91-94 (2009).
8. Kang, L., Wang, J., Zhang, Y., Kou, Z. & Gao, S. iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* **5**, 135-138 (2009).
9. Soldner, F. et al. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964-977 (2009).
10. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920 (2007).
11. Wernig, M. et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-324 (2007).
12. Carey, B.W. et al. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc Natl Acad Sci U S A* **106**, 157-162 (2009).
13. Carey, B.W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat Methods* **7**, 56-59.
14. Eggan, K. & Jaenisch, R. Differentiation of F1 embryonic stem cells into viable male and female mice by tetraploid embryo complementation. *Methods Enzymol* **365**, 25-39 (2003).
15. Eakin, G.S., Hadjantonakis, A.K., Papaioannou, V.E. & Behringer, R.R. Developmental potential and behavior of tetraploid cells in the mouse embryo. *Dev Biol* **288**, 150-159 (2005).
16. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat Methods* **7**, 53-55.
17. Lin, S.P. et al. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet* **35**, 97-102 (2003).
18. Zhou, Y. et al. Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. *Development* **137**, 2643-2652.
19. Hussein, S.M. et al. Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58-62.
20. Gore, A. et al. Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63-67.
21. Lister, R. et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*.
22. Doi, A. et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**, 1350-1353 (2009).
23. Markoulaki, S., Meissner, A. & Jaenisch, R. Somatic cell nuclear transfer and derivation of embryonic stem cells in the mouse. *Methods* **45**, 101-114 (2008).

24. Takada, S. et al. Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. *Hum Mol Genet* **11**, 77-86 (2002).

**ACKNOWLEDGEMENTS**:

We thank R. Flannery for help with mouse husbandry, D. Fu and members of the Jaenisch lab for excellent assistance and helpful comments. R.J. is supported by grants from the NIH: 5-RO1-HDO45022, 5-R37-CA084198, and 5-RO1-CA087869. J.H. is a Novartis Fellow by the Helen Hay Whitney Foundation.

**FIGURE LEGENDS**

**FIGURE 1: Generation of Col1a1 2lox iPSCs. (a)** Experimental scheme describing the Col1a1 2lox OSKM-Jae mouse strain for deriving iPSCs. **(b)** Immunostaining for alkaline phosphatase (Alk. Phos.), SSEA1 and Nanog on iPSC lines derived from four adult tissues (designation of donor tissue and cell line on top). **(c)** qRT-PCR of pluripotency genes in Col1a1 2lox OSKM-Jae iPSCs and control ES cell lines. Relative expression levels normalized to the housekeeping gene HPRT. **(d)** Hematoxylin and eosin (H&E) staining of teratomas induced after subcutaneous injection of Col1a1 2lox OSKM-Jae iPSCs into SCID mice indicates contribution to all three germ layers.

**(e)** Southern analysis of iPSCs obtained from sorted (GFP+) single iPSCs after transduction with adenovirus carrying Cre-recombinase. DNA was digested with XbaI and probed for correct targeting using a 5' Col1a1 probe. All lines showed band size (~ 4.7kb) expected for an untargeted Col1a1 locus [13] and lack of signal with an internal probe using Klf4. Same indicated iPSC lines are loaded in +Cre OSKM excision (right). **(f)** PCR genotyping of Col1a1 2lox OSKM-Jae iPSCs pre- and post-excision of reprogramming factor cassette. Homozygous targeted Col1a1 allele give rise to a band size ~ 550bp whereas untargeted allele is ~ 300bp. 1lox-Col1a1 after excision shows a slight higher band ~ 350-400bp. Ad, adenovirus.

**FIGURE 2: Generation of iPSC mice from Col1a1-factor-free iPSCs. (a)** Results summarizing the tetraploid (4n) complementation assay using two ES cell lines (Col1a1 2lox OSKM-Jae #1B,C ) and 9 Col1a1-factor-free iPSC lines. All data was generated 19.5 d.p.c.

Grey indicates visible re-absorptions (negative values), green indicates stillborn embryos with estimated development from E8-20, blue indicates live born pups that began breathing. Names of each cell line, total number of blastocysts transferred, and % found breathing (relative to blastocysts transferred) at birth are indicated below the graph. AVG column (far right) displays averages for all categories: "ESC" is from lines #1B and 1C and "iPSC" from OSKM-Jae derived iPSCs (both factor-free and factor-containing). P-value, student's t-test = 0.3071 for live born category **(b)** Representative brightfield images of nine pups from BC2 Ad.5 factor-free iPSC line (left panel). All are male with agouti fur. Two panels below show 4n mice from three other Col1a1-factor-free iPSC lines. **(c)** DNA titration curve for PCR genotyping assay demonstrate ~ 1-5% detection limit. **(d)** PCR assay for tetraploid cells and iPSC contribution to iPSC mice. Band size distinguishes 1lox-iPSCs from tetraploid host cells (BDF1 lane) and Col1a1 factor-containing cells (#660) indicating 4n-mice were derived from 1lox iPSC lines only. DNA was isolated from live born pups HD (head) and TL (tail & limbs) and from adult mice DNA used was from tail tissue.

**FIGURE 3: Silencing of Dlk1-Dio3 locus is not a strict marker of reduced pluripotency.** **(a)** qRT-PCR of nine Col1a1 2lox OSKM-Jae iPSCs and one ES cell control (V6.5) of *Gtl2* and *Rian*. Expression levels are relative to GAPDH. "Gtl2-LOW" is represented in iPS lines with <50% expression of both genes (right). Error bars indicated standard error of the mean of three biological replicates (SEM). **(b)** Brightfield pictures of Col1a1 2lox OSKM-Jae chimeras from Gtl2-LOW iPSCs. iPSC line pBL4 contributed to the germline. **(c)** Methylation analysis of IG-DMR in Col1a1 2lox OSKM-Jae iPSC lines. Genomic DNA was isolated from V6.5 ESCs, parental targeted V6.5 Col1a1 2lox OSKM-Jae ESCs, and six Col1a1 2lox OSKM-Jae lines. All DNAs were digested with *StuI* in combination with either methylation-insensitive *MspI*, or methylation-sensitive *HpaII* or *HhaI*. The digested DNAs were analyzed by Southern hybridization with a probe to IG-DMR. Restriction enzymes: M, *MspI*; Hp, *HpaII*; Hh, *HhaI*. **(d)** Quantification of chimeras generated from Col1a1 2lox OSKM-Jae strain using agouti coat color. Open green diamonds, chimerism of adult mice from parental V6.5 Col1a1 2lox OSKM-Jae ESCs; closed green diamonds iPSCs with ESC-like expression of *Gtl2* & *Rian* (Gtl2-ON); brown closed diamonds iPSCs with repressed expression of *Gtl2* & *Rian* (Gtl2-LOW). **(e)** Average chimerism for 44 mice analyzed from Col1a1 2lox OSKM-Jae

iPSCs. Gtl2-ON (57%); Gtl2-LOW (50%). Error bars indicate standard deviations. P-value, student's t-test = 0.4743. **(f)** Brightfield image of live born all-iPSC mouse from the Gtl2-LOW iPSC line pBL4. **(g)** Methylation of IG-DMR (*StuI+HhaI*) in 4n-mice and embryos using Southern blot protocol described above. Only unmethylated bands were observed from embryos and mice from iPSC line pBL4. As a control DNA of a live born pup generated from the Livp2D3 iPSC line shows both methylated and unmethylated banding patterns. HD, head; TL, tail & limbs.

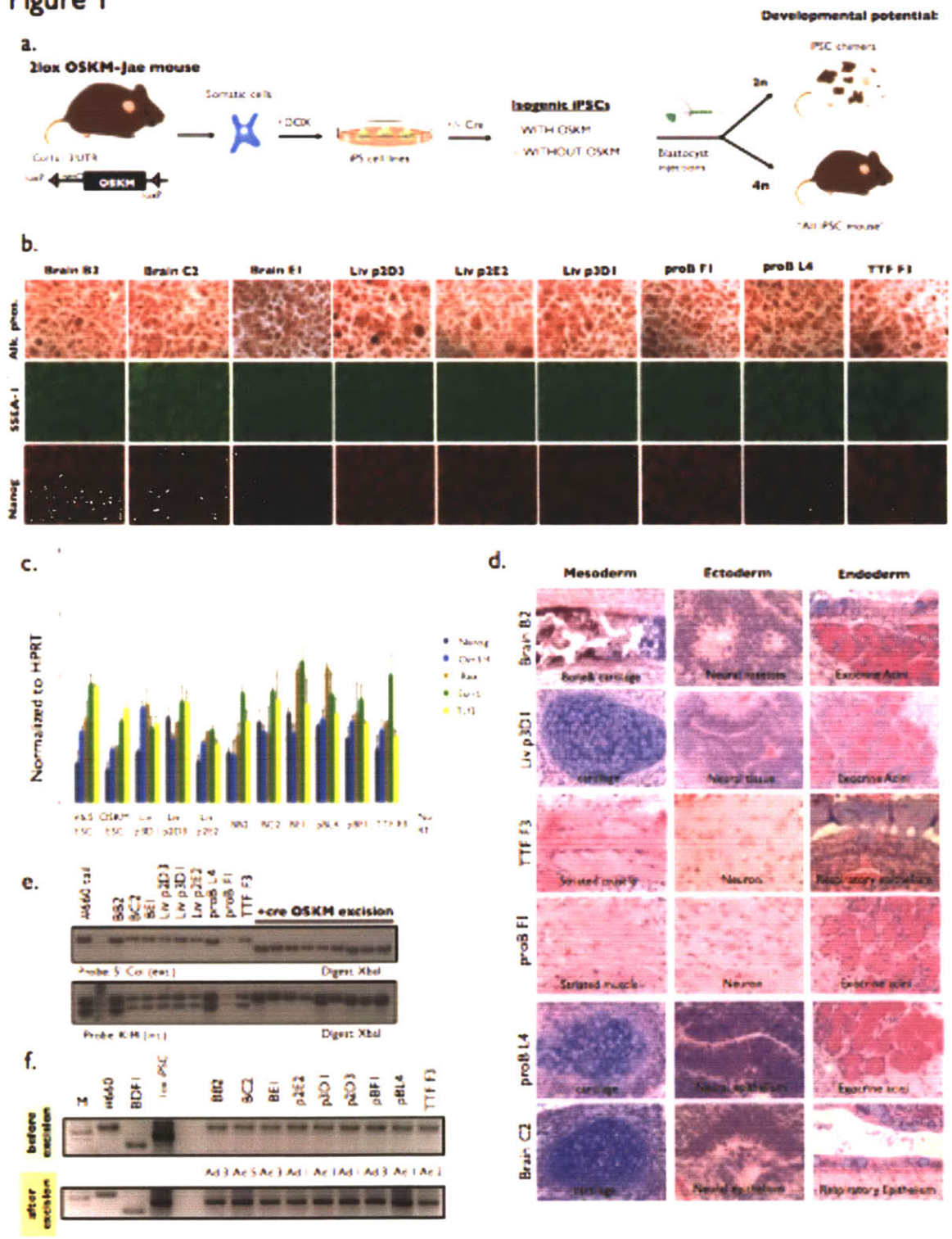
**FIGURE 4: Stoichiometry during reprogramming influences the epigenetic state of the Dlk1-Dio3 locus (a)** Quantitative RT-PCR in cells with DOX. 0.1, 2.0 is μg/ml of doxycycline, "_" is no DOX media. Error bars indicated standard deviation (S.D.) Values normalized to GAPDH and all samples were relative to rtTA only, which was set as "1." **(b)** Western blot analysis of reprogramming factor proteins upon induction with DOX in mouse embryonic fibroblasts used in (a). * indicates high molecular unprocessed precursor proteins for Oct4 and Klf4 **(c)** Quantification of protein levels detected by western blot. Images were processed using ImageJ (NIH). Reprogramming factor protein levels are shown as relative to actin for each sample (OSKM-Jae in green and OKSM-KH in blue). **(d)** Relative fraction of iPSCs recovered with aberrant DNA methylation as tested by Southern at the IG-DMR region (*StuI+HpaII* digest). Secondary OSKM-Jae MEFs were reprogrammed with no additional factor (-), additional Oct4, Klf4, or both (+Oct4+Klf4). Total number of iPSCs recovered and examined in each condition indicated below. P-value = fisher's exact T-test. **(e)** Representative southern blot analysis of IG-DMR (*StuI+HpaII* digest) from OSKM-Jae+Klf4 iPSCs generated from MEFs. iPSC lines with aberrant methylation (bold #) are indicated by loss of both lower molecular weight bands at ~ 250bp & 750bp. **(f)** Representative southern blot analysis of IG-DMR (*StuI+HpaII* digest) from iPSCs generated using OSKM-Jae+Oct4+Klf4. Bold # samples indicated iPSC lines with aberrant methylation at the IG-DMR locus. In some samples complete loss of bands was not observed due to higher presence of feeder MEFs.

**FIGURE 5: iPSC mice do not show elevated mortality from tumors**

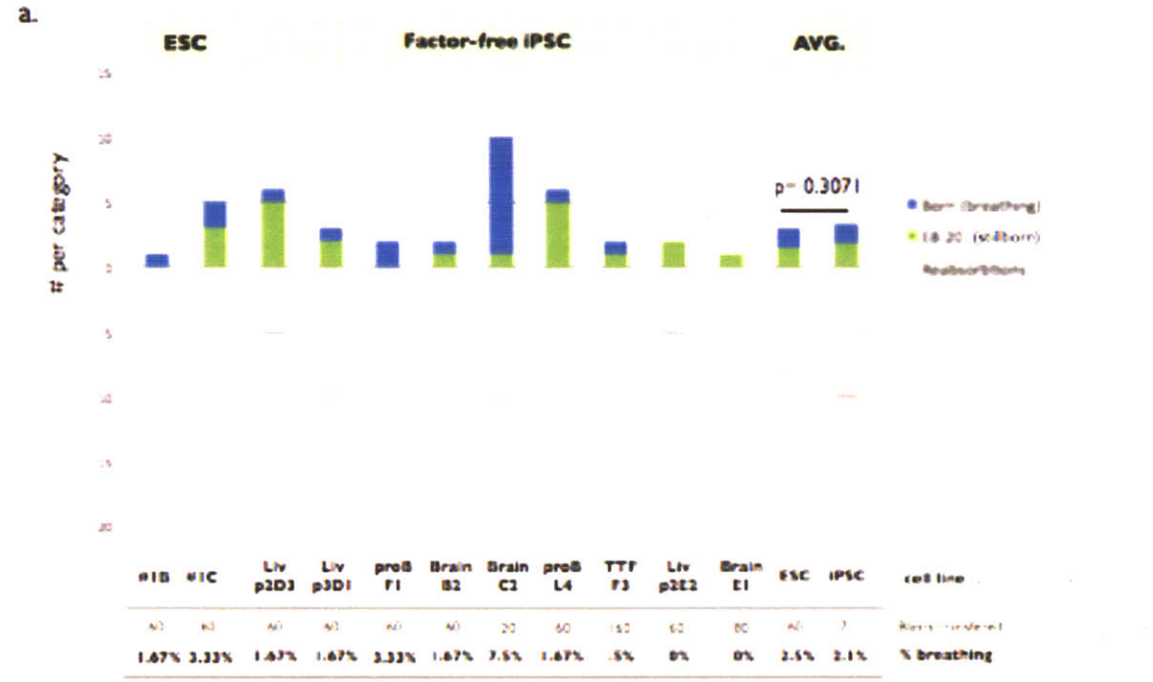
**(a)** Tumorigenicity and mortality of iPSC-derived mice from tumors. Orange line indicates

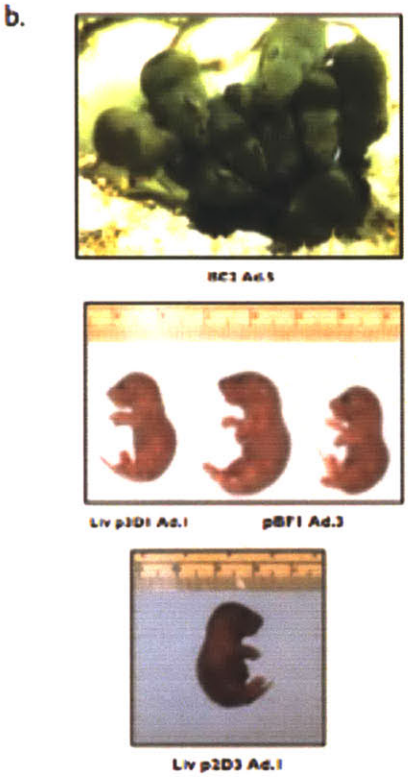
Col1a1 OSKM-Jae strain containing reprogramming factor transgene. Blue line indicates chimeric mice generated from factor-free OSKM-Jae iPSCs. **(b)** The cumulative mortality of Col1a1 OSKM-Jae iPSC-derived mice. Green line indicates F1 mice generated during maintenance of Col1a1 OSKM-Jae mouse colony. All sacrificed mice irrespective of cohort were dissected to evaluate the cause of death. Shown at the bottom of each figure are the numbers of mice analyzed at each time point.

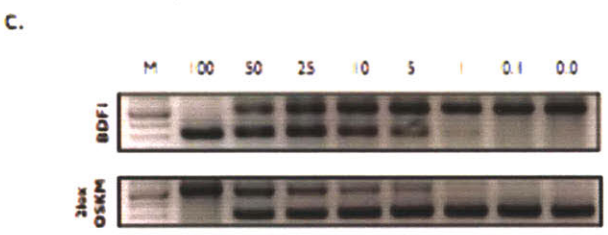
Figure 1

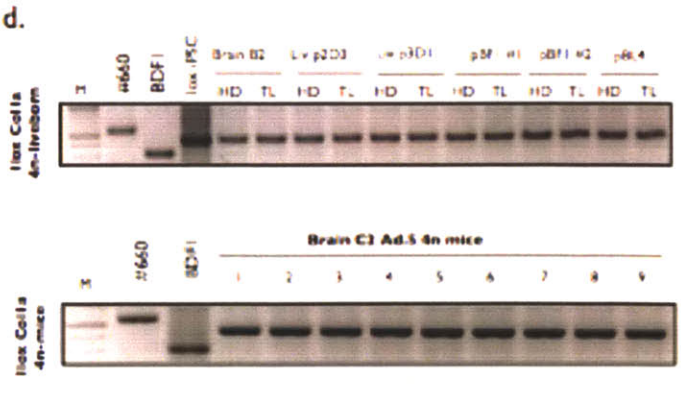

a.
2lox OSKM-Jae mouse
Somatic cells
+DOX
iPS cell lines
+/- Cre
Isogenic iPSCs
WITH OSKM
WITHOUT OSKM
Blastocyst injections
2n
4n
Developmental potential:
iPSC chimera
"All iPSC mouse"
OSKM
b.
Brain B2
Brain C2
Brain E1
Liv p2D3
Liv p2E2
Liv p3D1
proB F1
proB L4
TTF F3
Alk. phos.
SSEA-1
Nanog
c.
Normalized to HPRT
d.
Mesoderm
Ectoderm
Endoderm
Brain B2
Liv p3D1
TTF F3
proB F1
proB L4
Brain C2
e.
+cre OSKM excision
Probe 5' Cre (ext.)
Digest XbaI
Probe KM (int.)
Digest XbaI
f.
before excision
after excision

## Figure 2

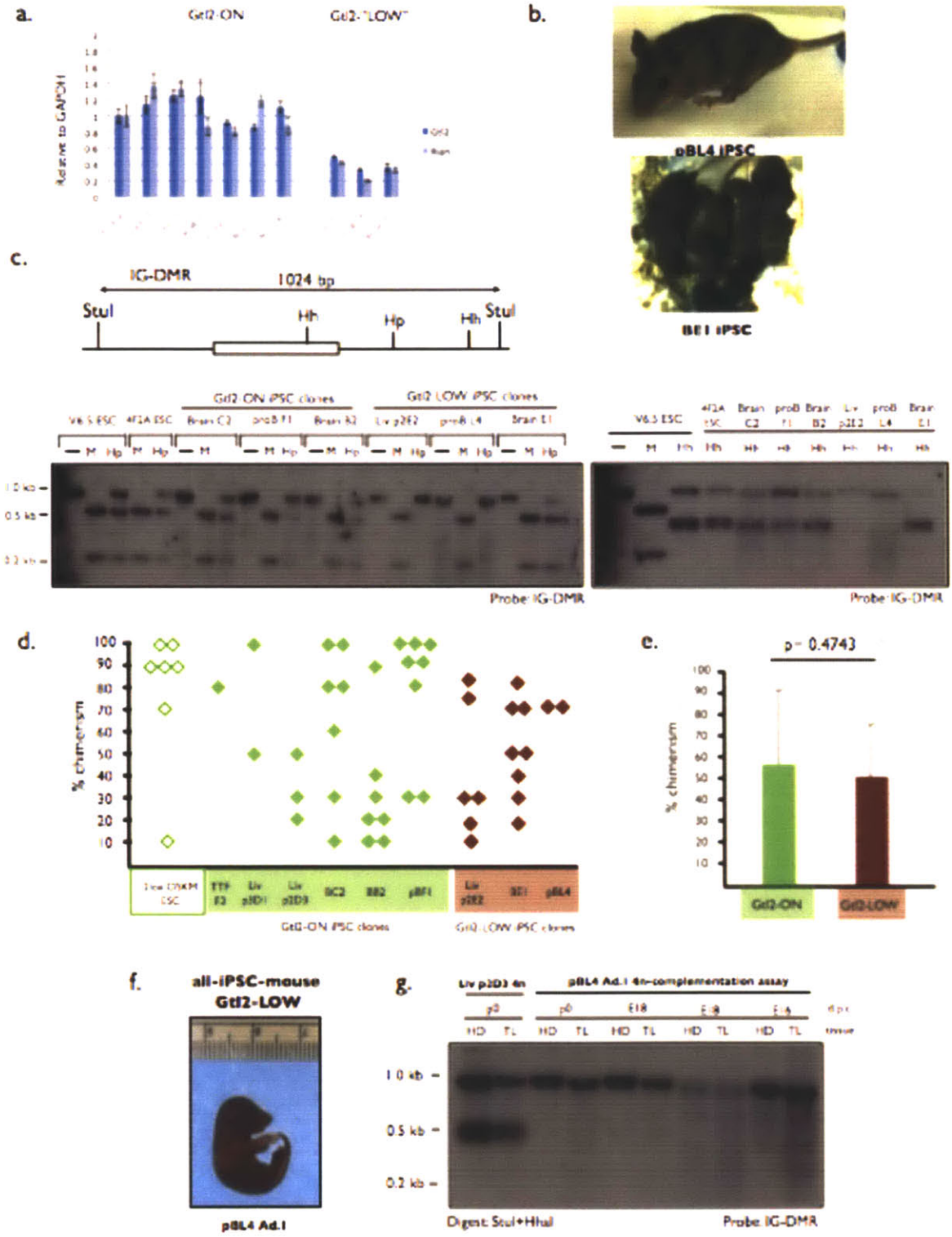

a.
ESC
Factor-free iPSC
AVG.
# per category
p= 0.3071
#1B
#1C
Liv p2D3
Liv p3D1
proB F1
Brain B2
Brain C2
proB L4
TTF F3
Liv p2E2
Brain E1
ESC
iPSC
cell line
% breathing


b.
BC1 Ad.5
Liv p3D1 Ad.1
pBF1 Ad.3
Liv p2D3 Ad.1

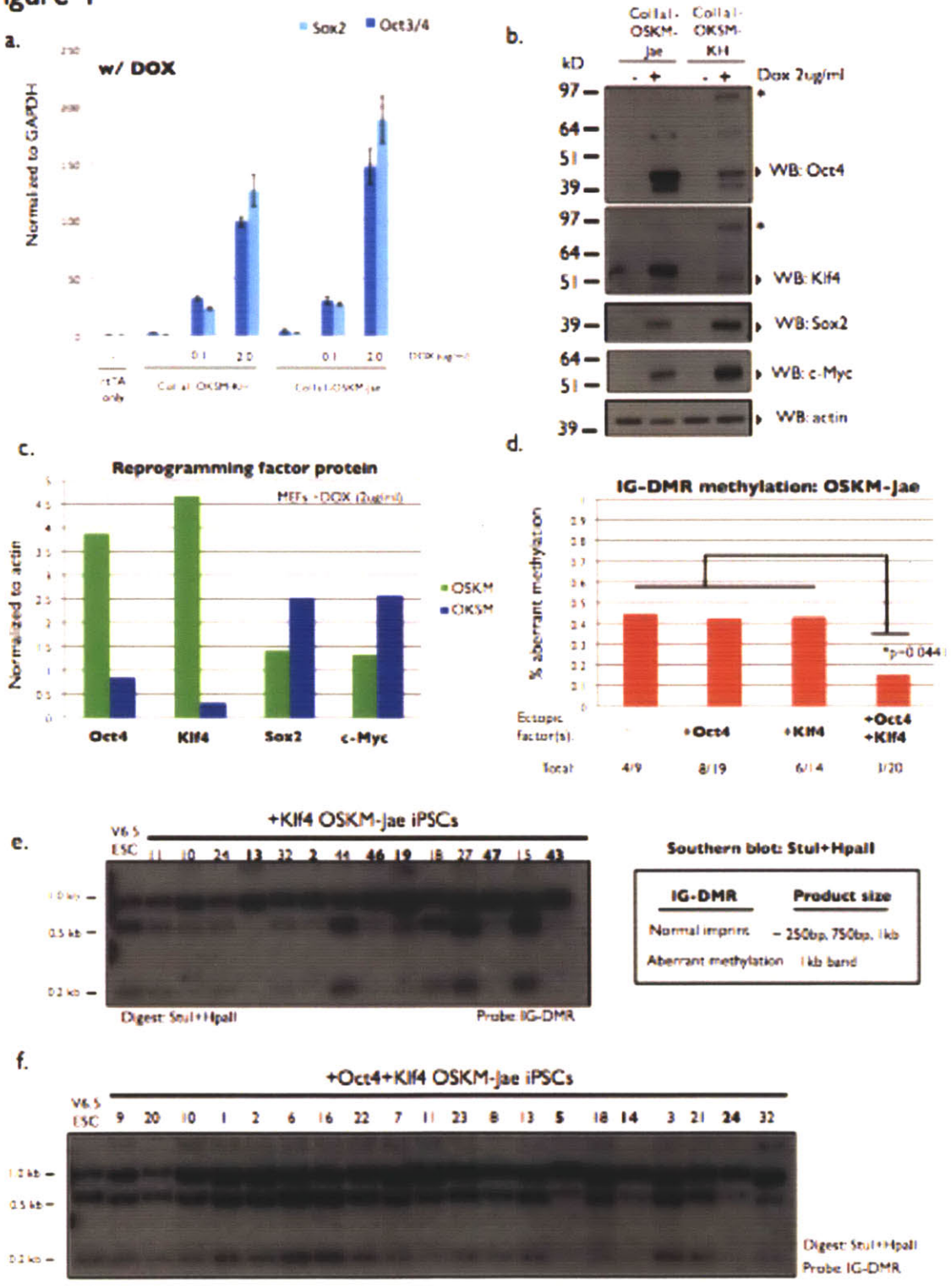

c.
M
100
50
25
10
5
1
0.1
0.0
BDF1
3lox OSKM


d.
M
#660
BDF1
Brain B2
Liv p2D3
HD
TL
3lox Cells 4n-liveborn
Brain C2 Ad.5 4n mice
3lox Cells 4n-mice

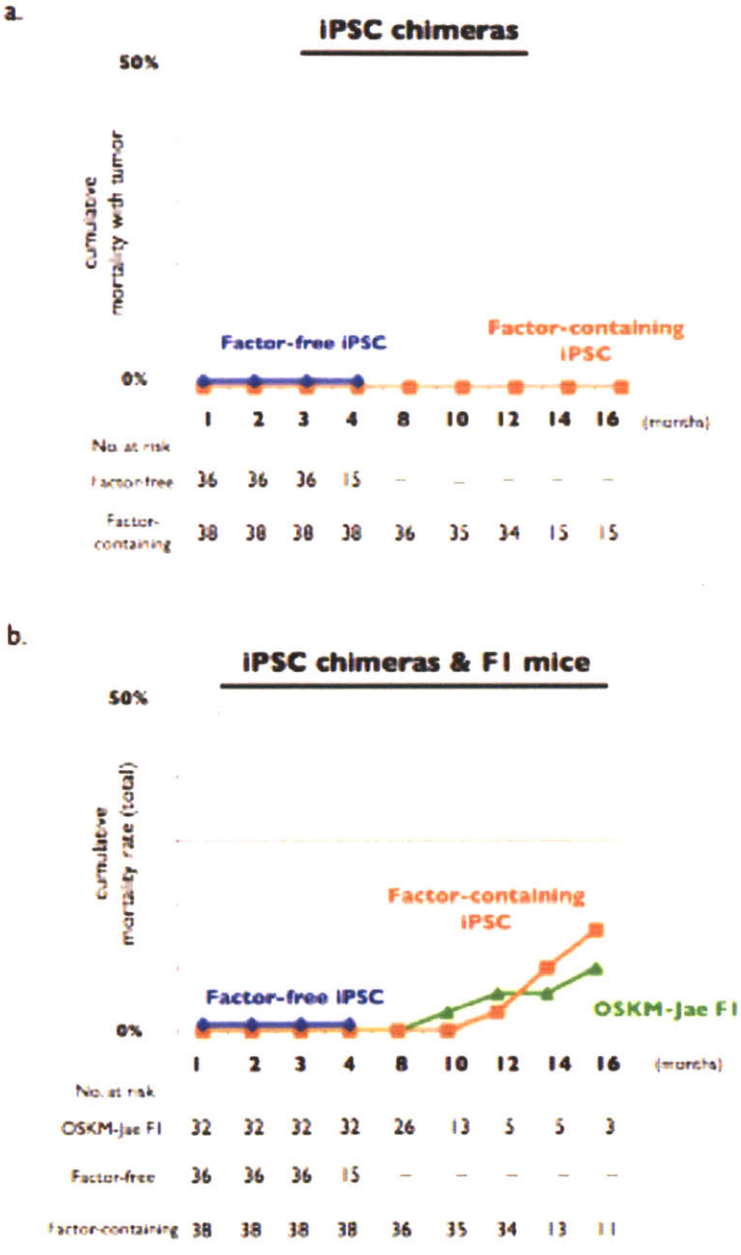
## Figure 3


a.
Gtl2-ON
Gtl2-"LOW"
Relative to GAPDH
b.
pBL4 iPSC
BE1 iPSC
c.
IG-DMR
1024 bp
StuI
Hh
Hp
Hh
StuI
Gtl2 ON iPSC clones
Gtl2 LOW iPSC clones
V6.5 ESC
4F2A ESC
Brain C2
proB F1
Brain B2
Liv p2E2
proB L4
Brain E1
M
Hp
1.0 kb
0.5 kb
0.2 kb
Probe: IG-DMR
V6.5 ESC
4F2A ESC
Brain C2
proB F1
Brain B2
Liv p2E2
proB L4
Brain E1
Hh
Probe: IG-DMR
d.
% chimerism
100
90
80
70
60
50
40
30
20
10
Gtl2-ON iPSC clones
Gtl2-LOW iPSC clones
e.
p= 0.4743
% chimerism
Gtl2-ON
Gtl2-LOW
f.
all-iPSC-mouse
Gtl2-LOW
pBL4 Ad.1
g.
Liv p2D3 4n
pBL4 Ad.1 4n-complementation assay
p0
E18
dpc
HD
TL
tissue
1.0 kb
0.5 kb
0.2 kb
Digest: StuI+HhaI
Probe: IG-DMR

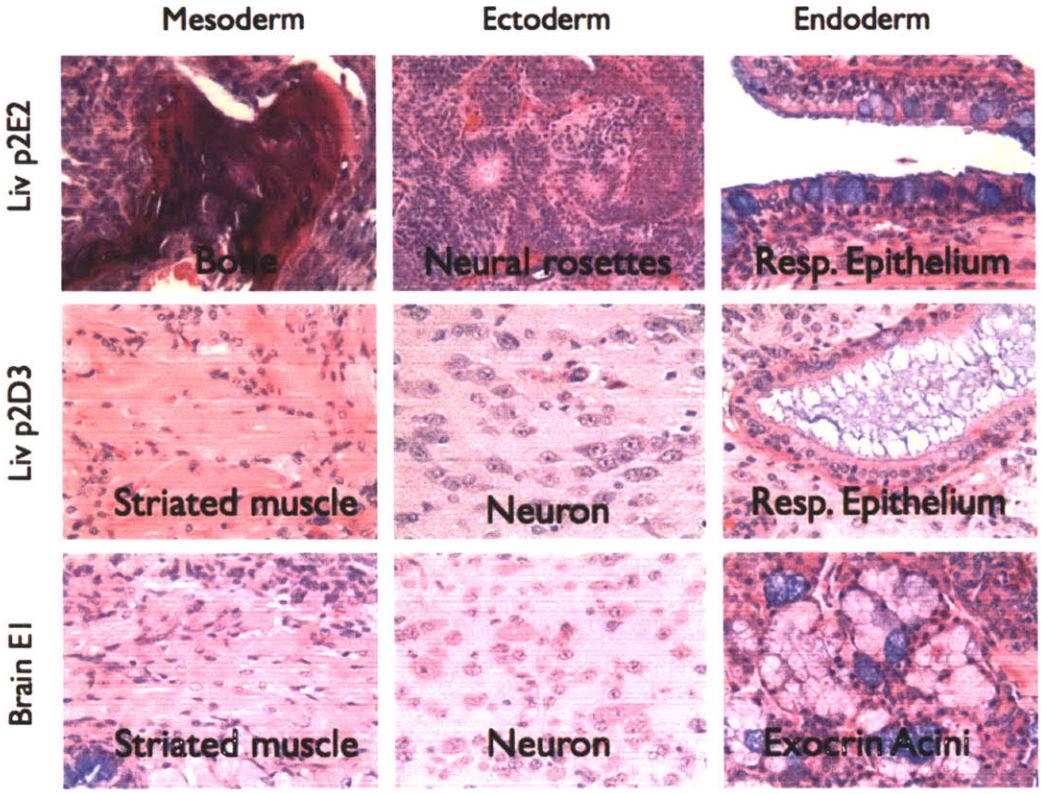
## Figure 4


a.
Sox2
Oct3/4
w/ DOX
Normalized to GAPDH
b.
Col1a1-OSKM-Jae
Col1a1-OKSM-KH
kD
Dox 2ug/ml
WB: Oct4
WB: Klf4
WB: Sox2
WB: c-Myc
WB: actin
c.
Reprogramming factor protein
MEFs +DOX (2ug/ml)
Normalized to actin
OSKM
OKSM
Oct4
Klf4
Sox2
c-Myc
d.
IG-DMR methylation: OSKM-Jae
% aberrant methylation
*p=0.044
Ectopic factor(s)
+Oct4
+Klf4
+Oct4 +Klf4
Total
4/9
8/19
6/14
1/20
e.
+Klf4 OSKM-Jae iPSCs
V6.5 ESC
11 10 24 13 32 2 44 46 19 18 27 47 15 43
1.0 kb
0.5 kb
0.2 kb
Digest: StuI+HpaII
Probe: IG-DMR
Southern blot: StuI+HpaII
IG-DMR
Product size
Normal imprint
~ 250bp, 750bp, 1kb
Aberrant methylation
1kb band
f.
+Oct4+Klf4 OSKM-Jae iPSCs
V6.5 ESC
9 20 10 1 2 6 16 22 7 11 23 8 13 5 18 14 3 21 24 32
1.0 kb
0.5 kb
0.2 kb
Digest: StuI+HpaII
Probe: IG-DMR

## Figure 5

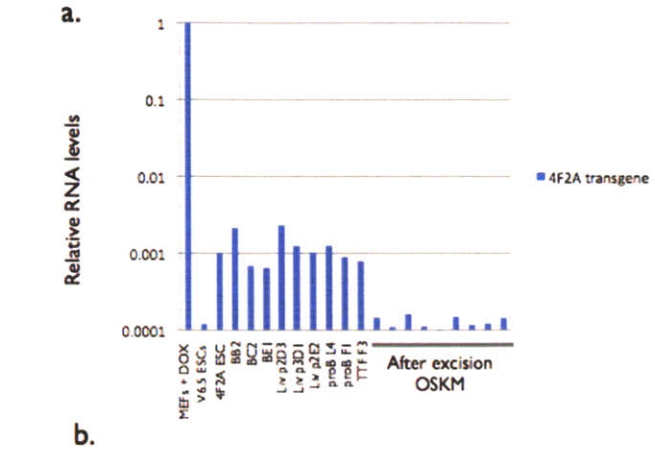

a.
iPSC chimeras
50%
cumulative mortality with tumor
Factor-free iPSC
Factor-containing iPSC
0%
1 2 3 4 8 10 12 14 16 (months)
No. at risk
Factor-free 36 36 36 15 – – – – –
Factor-containing 38 38 38 38 36 35 34 15 15
b.
iPSC chimeras & F1 mice
50%
cumulative mortality rate (total)
Factor-containing iPSC
Factor-free iPSC
OSKM-Jae F1
0%
1 2 3 4 8 10 12 14 16 (months)
No. at risk
OSKM-Jae F1 32 32 32 32 26 13 5 5 3
Factor-free 36 36 36 15 – – – – –
Factor-containing 38 38 38 38 36 35 34 13 11

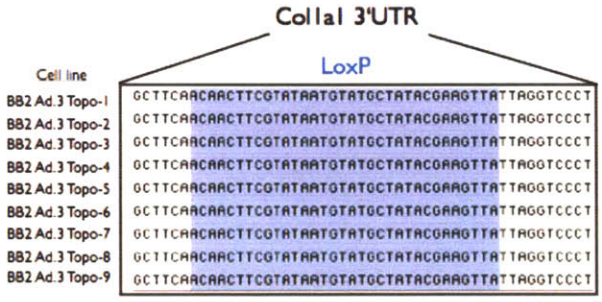
## Supplementary Figure 1




**Supplementary Figure 1**: Hematoxylin and eosin (H&E) staining of teratomas induced after subcutaneous injection of Col1a1 2lox 4F2A iPSCs into SCID mice indicates contribution to all three germ layers.

## Supplementary Figure 2

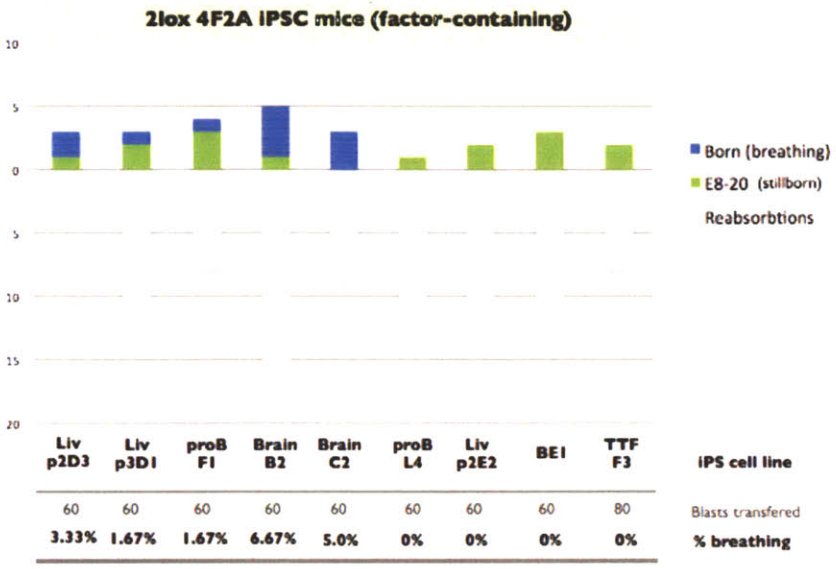


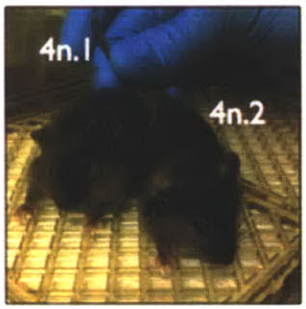

**Supplementary Figure 2:** (a) Quantitative RT-PCR of Col1a1 4F2A iPSCs using transgene-specific primers (E2A-cMyc see Materials and Methods) to detect residual transgene expression of 4F2A. Relative expression levels are normalized to DOX-induced expression in Col1a1 4F2A fibroblasts carrying single copy of rtTA and 4F2A. (b) Sequencing of Col1a1-1lox PCR band. Following isolation and gel purification DNA was cloned into TOPO2.1 vector (Invitrogen) following manufacturer's protocol. After transformation colonies were picked and plasmid DNA isolated. Sequencing was performed at Koch Center for Integrative Cancer Research. A single loxP site was found to be present in this band indicating the origin of the iPSCs from Col1a1 2lox 4F2A mouse strain.

# Supplementary Figure 3

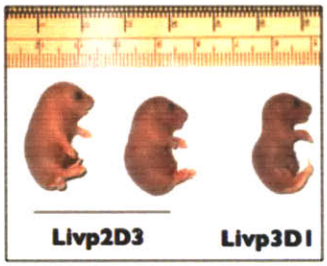
a.

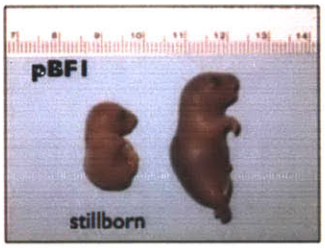



b.



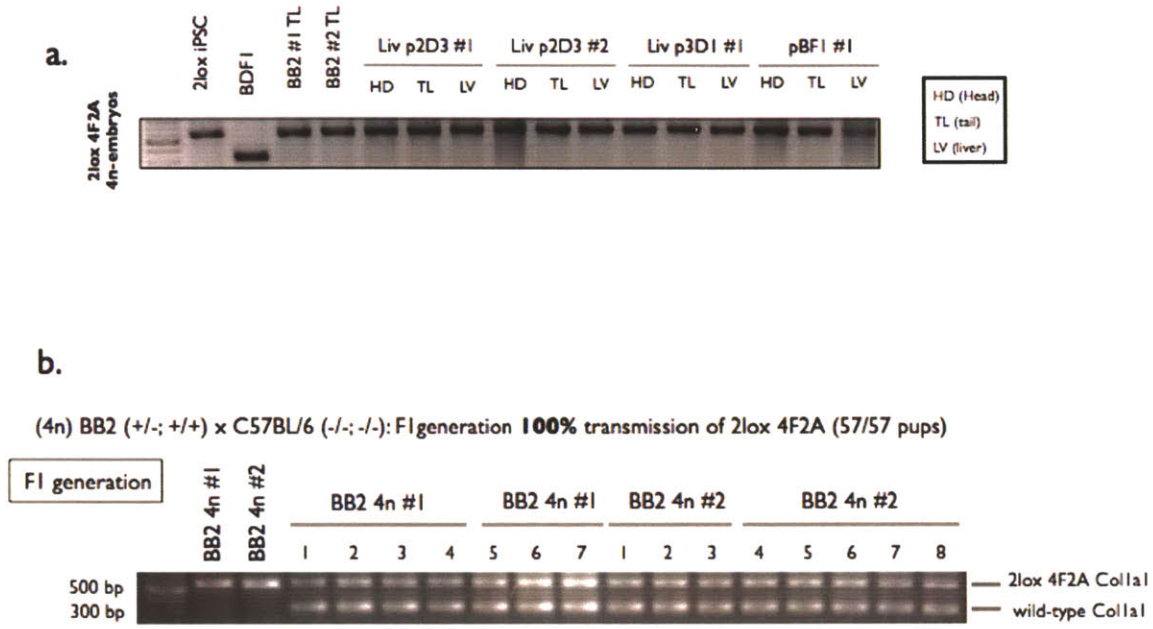

BB2 iPSC




**Supplementary Figure 3:** (a) Results summarizing the tetraploid (4n) complementation assay using 9 factor-containing Col1a1 2lox 4F2A iPSC lines. All data was generated 19.5 d.p.c. Grey indicates visible reabsorbtions (negative values), green indicates stillborn embryos with estimated development from E8-20, blue indicates liveborn pups that began breathing. Names of each cell line, total number of blastocysts transferred, and % found breathing (relative to blastocysts transferred) at birth are indicated below the graph. (b) Brighfield images of adult mice generated with BB2 iPSC and two additional lines. The liveborn pBF1 pup in the right panel shows enlargement not unexpected with tetraploid embryos.

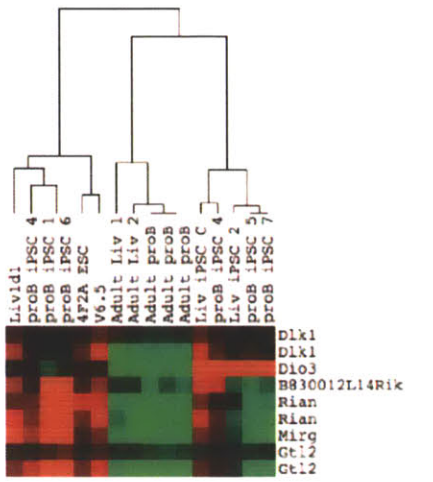
# Supplementary Figure 4




**Supplementary Figure 4:** (a) PCR assay for tetraploid cells and iPSC contribution to all iPSC mice. Band size indicates all mice were generated from Col1a1 2lox 4F2A iPSCs by only the upper (Col1a1 targeted) PCR band is visible. Tetraploid host cells (BDF1 lane) and DNA from TTF F3 iPSC. DNA was isolated from liveborn pups HD (head), TL (tail & limbs) and internal organs LV (liver). (b) Representative fenotyping of F1 pups from BB2 iPSC mice mated to BDF1 females. 100% germline transmission of one Col1a1 4F2A allele has been found in 57/57 offspring tested.

# Supplementary Figure 5

a.

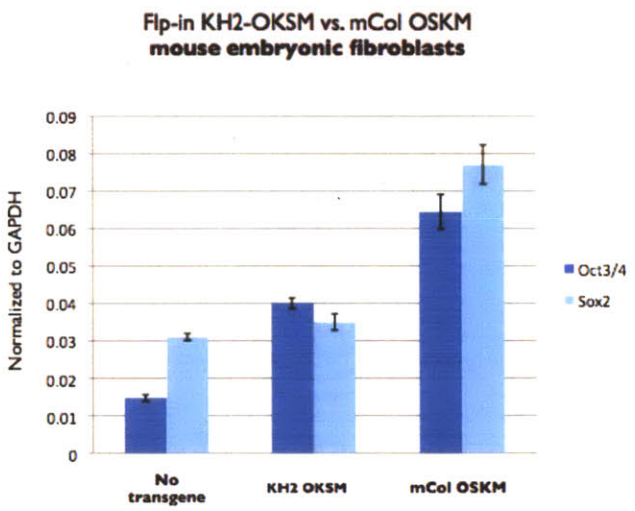



**6 out of 9** Col1a1 4F2A iPSCs show ESC-like expression of transcripts within the Dlk1-Dio3 imprinted cluster

b.

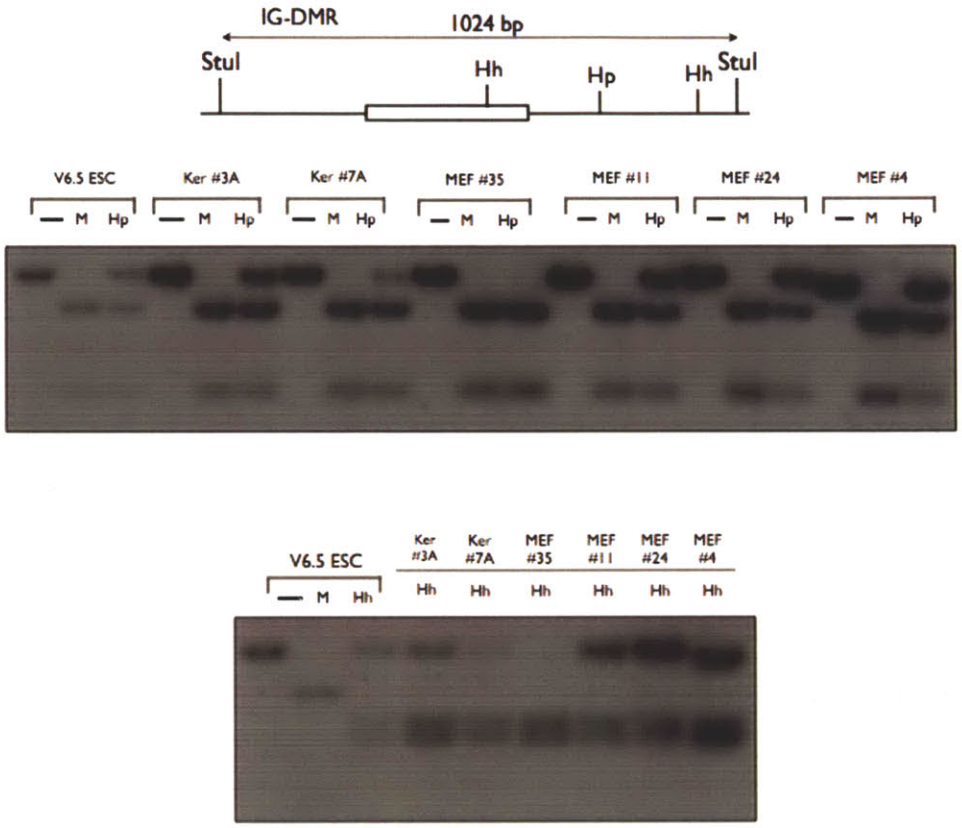
| | R26 M2rtTA | Col1a1 4F2A | Sex |
|---|---|---|---|
| 4F2A ESC | het | het | M |
| V6.5 ESC | - | - | M |
| Adult liver #291 | het | het | M |
| Adult liver #349 | het | het | M |
| Liver iPSC #1d1 | homo | het | F |
| Liver iPSC #2 | het | homo | M |
| Liver iPSC #C | het | homo | M |
| Adult proB #311 | homo | het | M |
| Adult proB #366 | het | het | M |
| Adult proB #409 | het | homo | M |
| proB iPSC #6 | homo | homo | M |
| proB iPSC #4 | homo | homo | M |
| proB iPSC #1 | homo | homo | M |
| proB iPSC #4A | het | homo | M |
| proB iPSC #7A | het | homo | M |
| proB iPSC #5 | het | homo | M |

c.




**Supplementary Figure 5:** (a) Hierachical clustering of microarray gene expression data using Cluster3 module with tissue-specific iPSCs from non-excisable Col1a1 4F2A iPSCs of liver and proB cell origins. Seven transcripts from within the Dlk1-Dio3 locus are indicated. Red, postive; Green, negative. Expression values are gene normalized for the whole array. Heatmaps were visualized in JavaTreeView. (b) Table indicating genotype and sex of iPSCs used in for microarrays. (c) Quantitative RT-PCR of uninduced transgenice MEFs from two reprogrammable mouse strains carrying single copies of rtTA and respective Col1a1 reprogramming vector (Carey et al., 2010 and Stadtfeld et al., 2010). Relative expression is normalized to GAPDH.

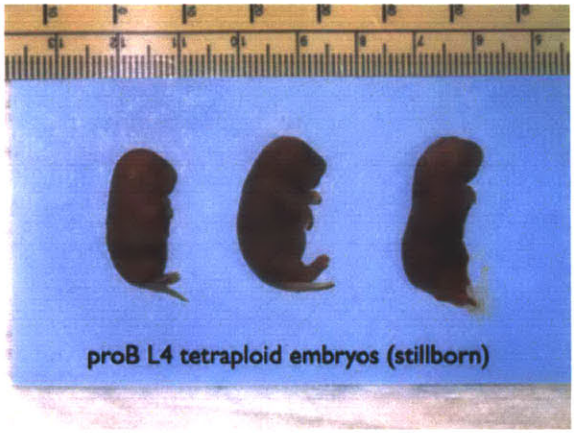
## Supplementary Figure 6



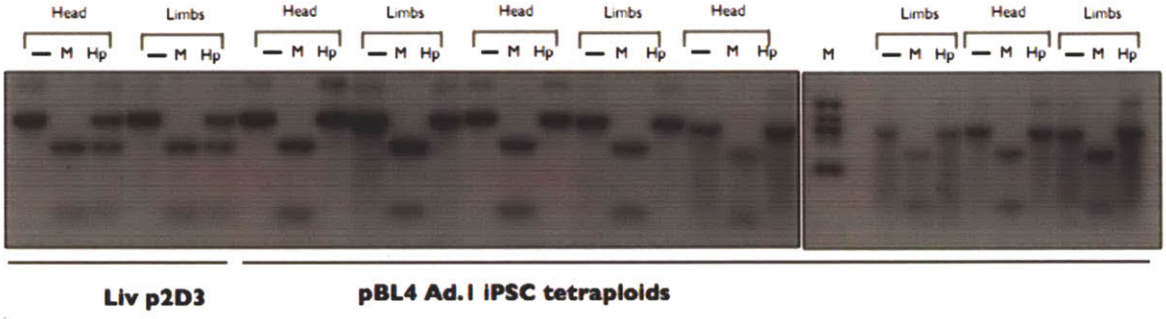

**Supplementary Figure 6:** Methylation analysis of IG-DMR in Col1a1 4F2A iPSC lines (non-excisable) from Keratinocytes and MEFs. All DNA was digested with StuI in combination with either methylation-insensitive MspI, or methylation-sensitive HpaII or HhaI. Images show data from Southern hybridization with a probe to IG-DMR. Restriction enzymes: M, MspI; Hp, HpaII; HH, HhaI.

## Supplementary Figure 7

a.




b.




**Supplementary Figure 7:** (a) Brightfield images of 4n-embryos from Col1a1-factor-free iPSC line pBL4 indicating late state development (~E16-18). (b) Methylation analysis of IG-DMR in 4n-embryos and liveborn pup. All DNA was digested with StuI in combination with either methylation-insensitive MspI, or methylation-sensitive HpaII or HhaI. The digested DNAs were analyzed by Southern hybridization with a probe to IG-DMR. Restriction enzymes: M, MspI; Hp, HpaII; HH, HhaI.

## Table 1: Summary of blastocyst injections

| | 4n injections (w/ OSKM) | | | 4n injections (w/o OSKM) | | |
|---|---|---|---|---|---|---|
| Cell line | Blastocysts Inj. (4n) | Embryos arrested at E10-E19.5 | Live pups (% inj.) | Blastocysts Inj. (4n) | Embryos arrested at E10-E19.5 | Live pups (% inj.) |
| BB2 | 60 | 1 | 4 (6.7) | 60 | 1 | 1 (1.7) |
| BC2 | 60 | 0 | 3 (5.0) | 120 | 1 | 9 (7.5) |
| BE1 | 60 | 3 | 0 | 80 | 1 | 0 |
| Liv p2D3 | 60 | 1 | 2 (3.3) | 60 | 5 | 1 (1.7) |
| Liv p3D1 | 60 | 2 | 1 (1.7) | 60 | 2 | 1 (1.7) |
| Liv p2E2 | 60 | 2 | 0 | 60 | 2 | 0 |
| pBL4 | 60 | 1 | 0 | 60 | 5 | 1 (1.7) |
| pBF1 | 60 | 3 | 1 (1.7) | 60 | 0 | 2 (3.3) |
| TTF F3 | 80 | 2 | 0 | 160 | 1 | 1 (0.5) |
| Total | 560 | | 11 (2.0) | 720 | | 16 (2.2) |

# Chapter 5

# CONCLUSIONS

In vitro reprogramming to pluripotency established an uncontroversial, technically unchallenging, practical system to generate patient-specific stem cells. Chapter 2 focused on addressing a more technical barrier that related to the use of multiple, independent viruses to deliver reprogramming factors to somatic cells. By designing novel vectors and adapting "2A" self-cleaving peptides we constructed a single virus capable of delivering up to four reprogramming factors (Oct3/4, Sox2, Klf4, and c-Myc) and successfully generating iPS cells in both mouse and human. Interestingly we found a reduced reprogramming efficiency, as indicated by reactivation of the endogenous Nanog locus, compared to delivery of reprogramming factors with single viruses. This suggests the reprogramming factor stochiometry using 2A peptides are distinct from that created by single viruses of all four reprogramming factors. Indeed, excess Sox2 during reprogramming inhibits generation of iPSCs (Eminli et al., 2008) from neural precursor cells. Other polycistronic cassettes that utilize a combination of "2A" peptides and internal ribosomal entry sites (IRES) have reported higher efficiencies of iPSC generation (Sommer et al., 2009) than that observed using "2A" peptides only. Next we found at least some iPS cell lines harbor only single integrations of the reprogramming provirus, suggesting reprogramming to pluripotency may be possible with minimal genetic modification to the genome. These results also established the possibility of excision of reprogramming factors upon isolation of iPS cells with one integration. To this end, two groups utilized the 2A-technology to generate single reprogramming cassettes capable of being excised upon establishment of iPS cells. Delivery by *piggbac* transposition established iPS cells with one integration that upon excision returned the integration site to a germline configuration (Woltjen et al., 2009), a similar report used simple mammalian expression vectors (Kaji et al., 2009). More recent work has now established that iPS

cells can be generated from human somatic cells by repeated delivery of reprogramming factor protein or RNA over a period of 2-6 weeks (Kim et al., 2009; Warren et al.) suggesting efficient generation of genetically unmodified hiPSCs will soon be a routine procedure.

Chapter 3 built on the previous chapter's observation that a single integration of our polycistronic cassette could generate iPSCs. Several technical barriers to studying reprogramming were addressed by "single-gene" reprogrammable mouse transgenic strains. The ability to maintain a mouse colony that contained drug-inducible reprogramming factors was not possible without concerns of tumors and laborious genotyping. Previous reprogrammable mouse strains, or secondary systems, utilized single viruses to establish clonal primary iPS cells lines that were injected into blastocyst embryos to generate somatic cells that harbored genetically identical proviral integration patterns. However these lines typically had >7 integrations, and to maintain a homozygous mouse colony required careful and extensive genotyping work with each F1 generation. Moreover many mice developed tumors limiting the use of these strains for long-term analysis (Markoulaki et al., 2009). Using a single vector carrying all four reprogramming factors we instead targeted ES cells that were then injected to create a mouse colony whereby four reprogramming factors were expressed from a single genomic locus (collagen 1a1). This not only afforded us the ability to maintain homozygous mouse colonies but also allowed easy transfer to different genetic backgrounds without concerns for tumors, which we do not observe in these strains. In addition we established that polycistronic vectors could potentially be used to screen for genetic or chemical compounds that replace individual factors in reprogramming, such as Sox2. Finally the ability to excise reprogramming factors upon establishment of iPSCs gener-

ated pluripotent cells that, other than a single loxP site within the 3'UTR of the Col1a1 locus, harbored genetically unmodified genomes.

Chapter 4 describes an extensive set of experiments on transgene-free iPSCs from distinct adult tissues, something not possible with any other reprogramming system published to date. These experiments attempted to address a nagging question and observation in the reprogramming field, that most iPSC lines created to date do *not* fulfill all developmental criteria associated with naive pluripotency in ESCs. First we extend the findings of others that report the successful generation of tetraploid-competent iPS cells (Boland et al., 2009; Kang et al., 2009; Stadtfeld et al.; Zhao et al., 2009). These authors report iPS cells from embryonic and adult mouse fibroblasts at efficiencies similar to ES cell lines (0.5-13%). However, the isolation of an 4n-competent iPS cell line during an individual reprogramming experiment was typically quite rare (3/37 (Zhao et al., 2009) and 4/61 (Stadtfeld et al.)). Whether these observations are technical and not biological differences between iPSCs and ESCs has yet to be addressed. Two studies analyzing gene expression differences between 4n-competent and 4n-non-competent iPSCs proposed silencing of a single imprinted gene locus at Chr 12qF1 was responsible for the reduced potency of some iPSC lines (Liu et al.; Stadtfeld et al.). Stadtfeld et al. (2009) report that the vast majority of lines have aberrantly silenced this locus (4/61, or ~ 6%). In contrast, we observe that 4F2A-induced iPSC lines do not generally silence the *Dlk1-Dio3* locus with ~ 66% of derived lines showing normal expression of genes within this imprinted locus. Whether transcriptional differences are underlying these disparate results upon generation of tissue-specific iPS cells were ruled out by analyzing transgenic MEFs from each strain, suggesting post-transcriptional differences generate qualitatively different iPSCs. Finally, we tested whether iPSCs from defined adult tis-

sues could fulfill all developmental criteria associated with ES cells, such as tetraploid complementation tests. We find the majority (~ 80%) of our iPSCs independent of this tissue-of-origin are tetraploid competent. Previous reports have identified differential methylation or altered histone modification patterns in iPSCs, few studies have addressed whether these are functionally relevant in differentiation towards somatic lineages in vivo.

## FUTURE APPLICATIONS OF iPSC TECHNOLOGY

Overcoming the major ethical and practical hurdles associated with embryonic stem cell research was achieved following the discovery of iPS cells. This technology not only affords a novel tool to dissect nuclear reprogramming in vitro, but will also help establish new in vitro models of many human diseases (Saha and Jaenisch, 2009). The vast majority of patients with neurological disorders such as Parkinson's or Alzheimer's disease do not harbor single mutations causing the disease, with most cases manifesting "sporadic," non-familial forms of onset and disease occurrence itself. While mouse models of many diseases do exist, frequently the disease phenotype only partially resembles the human pathophysiology. Issues arising from the inbred nature of mouse strains can also generate unreliable phenotypes. Therefore use of human patient-derived induced pluripotent stem cells may afford unique insights into disease pathologies by controlling for genetic background in which the disease occurs and generating the diseased cell type in vitro. These efforts will require genetic manipulations to hiPSCs, potentially using highly efficient gene targeting strategies such as zinc-finger nucleases (Hockemeyer et al., 2009) to generate reporter lines for candidate genes of the diseased cell type.

Human ES or iPS cells are unique when compared to blastocyst derived ES cells in mouse, and more likely reflect the post-implantation epiblast derived stem cells of mouse, termed EpiSCs (Tesar et al., 2007). These two pluripotent in vitro established cell types have been called "naive" for the ICM-derived mouse ES cells and "primed" for the post-implantation derived mouse EpiSCs. EpiSCs have unique gene expression profiles, epigenetic marks, X-chromosome states, and growth requirements when compared to ESCs and experiments using "2i" conditions in addition to other compounds established that further reprogramming of hESCs or hiPSCs can generate naive pluripotent cells with all the aforementioned molecular hallmarks of naive ES or iPS cells of mouse (Hanna et al. 2010). Although the naive pluripotent state in humans was shown to technically exist it was stabilized only transiently, suggesting further enhancements to the current protocol should yield more stable cells that reflect the "ground state" of pluripotency in mouse.

Induced pluripotency has also opened up new areas of exploration for nuclear reprogramming generally. It remains unclear how induced pluripotency occurs after expression of reprogramming factors and applications such as single cell analysis may offer novel insights into why some iPS cells appear earlier in time (Hanna et al., 2009). The role of each individual reprogramming factor also remains to be clarified, however many experiments point to Oct3/4 being one of the most crucial factors. Whether individual reprogramming factors can impart unique epigenetic or gene expression states following establishment of iPSCs also remains to be thoroughly investigated. The use of polycistronic cassettes may offer novel means to test whether additional expression of some factors may induce some of the molecular phenotypes described to date in iPSCs,

suggesting many observations so far can most likely be overcome with further optimization and study of the reprogramming process.

## Bibliography


Boland, M.J., Hazen, J.L., Nazor, K.L., Rodriguez, A.R., Gifford, W., Martin, G., Kupriyanov, S., and Baldwin, K.K. (2009). Adult mice generated from induced pluripotent stem cells. Nature *461*, 91-94.

Eminli, S., Utikal, J.S., Arnold, K., Jaenisch, R., and Hochedlinger, K. (2008). Reprogramming of Neural Progenitor Cells into iPS Cells in the Absence of Exogenous Sox2 Expression. Stem Cells.

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc Natl Acad Sci U S A *107*, 9222-9227.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595-601.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B., *et al.* (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol *27*, 851-857.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771-775.

Kang, L., Wang, J., Zhang, Y., Kou, Z., and Gao, S. (2009). iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. Cell Stem Cell *5*, 135-138.

Kim, D., Kim, C.H., Moon, J.I., Chung, Y.G., Chang, M.Y., Han, B.S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., *et al.* (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell *4*, 472-476.

Liu, L., Luo, G.Z., Yang, W., Zhao, X., Zheng, Q., Lv, Z., Li, W., Wu, H.J., Wang, L., Wang, X.J., *et al.* Activation of the imprinted Dlk1-Dio3 region correlates with pluripotency levels of mouse stem cells. J Biol Chem *285*, 19483-19490.

Markoulaki, S., Hanna, J., Beard, C., Carey, B.W., Cheng, A.W., Lengner, C.J., Dausman, J.A., Fu, D., Gao, Q., Wu, S., *et al.* (2009). Transgenic mice with defined combinations of drug-inducible reprogramming factors. Nat Biotechnol *27*, 169-171.

Saha, K., and Jaenisch, R. (2009). Technical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell *5*, 584-595.
Smith, Z.D., Nachman, I., Regev, A., and Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nat Biotechnol *28*, 521-526.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543-549.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175-181.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618-630.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., *et al.* (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766-770.

Zhao, X.Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.L., Ma, Q.W., Wang, L., *et al.* (2009). iPS cells produce viable mice through tetraploid complementation. Nature *461*, 86-90.

# APPENDIX I:

## Genome-wide maps of chromatin modifications in tissue-specific iPS cells

**NOTE:** This data was generated in close collaboration with Menno Creyghton and Albert Cheng. The data and analysis presented herein on adult tissues was used in Creyghton et al., 2009 PNAS 107(50):21931-6.

Reprogramming to pluripotency by defined factors is a robust protocol to generate pluripotent cells from adult somatic cells, yet variability in developmental potential between iPS cell lines has been reported by several groups [1,2]. Termed "epigenetic memory," some iPSC lines were shown to prefer to differentiate back into their tissue-of-origin when compared to iPSCs derived from other unrelated somatic cell lineages. Whether these observations are inherent to the induced pluripotent state or due to technical limitations of the current reprogramming technologies remains under investigation. We sought to address these questions using a reprogrammable mouse strain in which we isolated iPSCs from three separate lineages (e.g. CD19+ proB-cells and adult liver) and compared their molecular and developmental potential to ES cells. We focused on cell-type-specific cis-regulatory noncoding elements such as enhancers, which have been previously demonstrated to harbor unique chromatin signatures in addition to being highly-conserved elements [3,4].

We tested global gene expression by microarray analysis and also performed ChIP-Seq on various histone H3 tail modifications such as H3K4me3, H3K4me1, H3K27me3, and H3K27ac in tissue-specific iPSCs as well as the adult tissues from which they were derived. In adult tissues 94,437 different putative enhancer regions were identified on the basis of distal H3K4me1. Similar to previous reports, Creyghton et al. (2010) show H3K4me1 assci-

ated with a single cell type is also associated with gene activity in that specific cell type [5]. Motif search analysis on conserved DNA sequence identified DNA binding motifs for known tissue-specific enhancer regulators (e.g. FoxA2) [3-5] with integration of ChIP-Seq datasets for each transcription factor showing enrichment at the identified H3K4me1 region. Similar results were found using 40,274 enriched promoters distal regions for H3K27ac [5]. Interestingly, Creyghton et al. show enhancers that are H3K4me1+ but H3K27ac- are void of gene activity and suggest that H3K4me1+ enhancers lacking H3K27ac mark "poised" enhancer elements associated with potential future gene activity.

We employed similar analysis on tissue-specific iPSCs to determine the extent of reprogramming of histone marks associated with cell-type-specific enhancer elements. We find that the majority of tissue-specific iPSCs, independent of their tissue-of-origin, successfully reprogram global histone modification profiles to that of an ESC-like state. However some iPSC lines show regions of unchanged histone modifications (e.g. H3K4me1) associated with "inactive" cell-type-specific cis-regulatory elements. Interestingly, these lines exhibit developmental potential similar, but not identical to ESCs, in they can generate teratomas and liverborn chimeras. In agreement with the transcriptional analysis demonstrating these iPSCs are highly similar to ESCs, histone marks associated with "active" enhancers regulating gene activity (e.g. H3K27ac), show complete reprogramming to an ESC-like state. These data argue gene expression states do not necessarily reflect a fully reprogrammed epigenome. Successful reprogramming of certain histone modifications but not others, such as those associated with inactive or "poised" enhancers (e.g. H3K4me1), need not necessarily reprogram to an ESC-like state. Finally, these lines show incomplete reprogramming of H3K4me1 and H3K4me3 at key pluripotency genes such as Oct4 and Klf2,

suggesting that conclusions of "epigenetic memory" may not necessarily reflect an ESC-like state given "incomplete reprogramming" of pluripotency genes cannot be excluded as the cause of limited developmental potential of certain iPSC lines. These results are similar to previous reports in which incomplete reprogramming of the endogenous Nanog locus was observed in some iPSC lines that exhibited "epigenetic memory" phenotypes in vitro [1].

**MATERIALS AND METHODS** (from Creyghton et al., *PNAS* 2010):

**Cell Growth and Culture Conditions**. Blastocysts were isolated from the oviducts of hormone-primed [pregnant mare serum at 5 IU/ mouse (catalog no. 367222; Calbiochem), human chorionic gonadotrophin at 50 IU/mL (catalog no. 230734; Calbiochem)] B6D2F1 mice (Taconic) either 3 d after fertilization or 1 d after fertilization followed by 3 d of culture in potassium simplex optimized media (Specialty Media). C57/BL6-129JAE (v6.5) murine ES cells and J21 (C57/BL6) ES cells were grown in DMEM supplemented with 15% FBS (HyClone), 1,000 U/mL leukemia inhibitory factor 0.001% β-mercapto-ethanol (M7522; Sigma), 100 μM nonessential amino acids (11140-050; Invitrogen), 2 mM L-glutamine (25030- 081; Invitrogen), 100 U/mL penicillin, and 100 μg/mL streptomycin (15140-122; Invitrogen). Adult tissues were isolated from 4- to 6-wk-old 4f2a mice [6]. CD19+ proBcells were isolated by magnetic affinity cell sorting cell separation (Miltenyi Biotech) according to the manufacturer's protocol. Purified B cell subsets were resuspended in Iscove's Modified Dulbecco's Medium with 15%FCS as well as IL-4, IL-7, and stem cell factor (10 ng ml−1 each; Peprotech) and plated on OP9 bone marrow stromal cells (ATCC). Cells were fixed in formaldehyde or TRIzol reagent for ChIP and RNA isolation. For isolation of liver cells, mice were first perfused with 50 mL HBSS buffer (without Ca2+ and Mg2+), then with 50 mL HBSS (without Ca2+ and Mg2+) containing collagenase (type IV) (Sigma) (100 U mL−1). Liver was dissected away from surrounding tissues and dissociated in 10mLDAGmedia [phenol-red freeDMEM(Gibco) and BSA 1 g per 0.5 L] and filtered two times through a sterile 100-μm cell strainer. Liver cell preparations were centrifuged at 30 °— g for 3 min at 4 °C, and the cells were washed two times in PBS. RNA Iso-

lation and Microarray Analysis. RNA was isolated using TRIzol reagent (Invitrogen) according to the manufacturer's protocol and DNase treated using the DNA-Free RNA kit (R1028; Zymo Research). Cy3 dye-labeled cRNA samples wereprepared using Agilent's QuickAmp sample labeling kit. Input was 0.5 μg total RNA. Briefly, first- and second-strand cDNA are generated using Moloney Murine Leukemia Virus–reverse transcriptase enzyme and an oligo-dT based primer. In vitro transcription
is performed using T7 RNA polymerase and either cyanine 3-CTP or cyanine 5-CTP, creating a direct incorporation of dye into the cRNA. Agilent (mouse 4x44k) expression arrays were hybridized according to our laboratories method, which differs slightly from the Agilent standard hybridization protocol. The hybridization mixture consisted of 1.65 μg cy3 dye-labeled cRNA for each sample, Agilent hybridization blocking components, and fragmentation buffer. The hybridization mixtures were fragmented at 60 °C for 30 min, followed by addition of Agilent 2X hybridization. The arrays were then hybridized for 16 h at 60 °C in an Agilent rotor incubator set at maximum speed. Arrays were washed (6°— SSPE/0.005% N-laurylsarcosine) on a rotating platform at room temperature for 2 min, and then washed again (0.06°— SSPE) for 2 min at room temperature. The arrays were then dipped briefly in acetonitrile, followed by 30 s in Agilent Stabilization and Drying Solution. Arrays were scanned with an Agilent scanner using the Agilent feature extraction software.

**Chromatin Immune Precipitation.** Cells were chemically cross-linked eitheronplate or in suspension by the addition of one-tenth volume of fresh 11% formaldehyde solution containing [1mMEDTA, 0.5 mMEGTA, 100mMNaCl, and 50mMHepes-KOH (pH 7.5)] for 15 min at room temperature. Cells were rinsed twice with 1°— PBS, harvested using a silicon scraper or centrifuge, flash-frozen in
liquid nitrogen, and stored at −80 °C before use. Typically 1–3 °— 108 cells were resuspended in 10 mL lysis buffer [50 mM Hepes- KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Nonidet P-40, and 0.25% Triton X-100] and rocked at 4 °C for 10 min. Cells were pelleted at 2880 °— g on a tabletop centrifuge at 4 °C
and resuspended in wash buffer [200mMNaCl, 1mMEDTA, 0.5 mM EGTA, and 10 mM Tris (pH 8.0)] and rocked at room temperature for 10 min. Cells were pelleted by spinning at

2880 °— g/ 4 °C for 5 min and resuspended in 3 mLof sonication buffer [1mM EDTA, 0.5mMEGTA, 10mMTris (pH 8.0), 100mMNaCl, 0.1%
Na-deoxycholate, and 0.5% N-lauroylsarcosine]. We used a Misonix Sonicator 3000 and sonicated at ≈20 W for 8 °— 30-s pulses (60-s pause between pulses) at 4 °C for ES cells while samples were immersed in an ice bath. For adult tissues we sonicated six to seven rounds of 30-s pulses. Triton X-100 was added to the resulting whole-cell extract (1% end concentration), which was then cleared by centrifugation (20,800 °— g at 4 °C), and supernatant was incubated overnight at 4 °C with 100 μL of Dynal ProteinGmagnetic beads that had been preincubated with 10 μg of the appropriate antibody for at least 3 h. Beads were washed five times with RIPA
buffer [50mMHepes (pH 7.6), 1mMEDTA, 0.7% deoxycholate, 1% Nonidet P-40, and 0.5 M LiCl] and once with Tris/EDTA buffer (TE) containing 50 mM NaCl. Bound complexes were eluted from the beads by heating at 65 °C with occasional vortexing
in elution buffer [50mMTris (pH 8), 10mMEDTA, and1%SDS], and cross-linking was reversed by overnight incubation at 65 °C. After removal of the beads, immunoprecipitatedDNAwas diluted 1:1 with TE and then treated with RNaseA(0.2 μg/μL final) for 2 h at 37 °C, followed by proteinase K (0.2 μg/μL final) treatment for 2 h at 50 °C. DNA was purified using two consecutive phenol:chloroform extractions using Phase Lock Gel tubes (5′) and once using Qiagen PCR purification columns. The resulting DNA was either used for gene-specific PCR or further treated for analysis on the Solexa sequencer (GA2X genome sequencer) using the ChIP seq
sample prep kit (1003473; Illumina) according to the manufacturer's protocol (11257047; Illumina), selecting library fragments between 200 and 350 bp. Samples were run by the Massachusetts Institute of Technology biopolymers facility
(http://web.mit.edu/ki/facilities/biopolymers/index.html) using the GA2X genome sequencer (SCS v2.6, pipeline 1.5).

**Data Analysis.** Images that were acquired from the Illumina/Solexa sequencer were processed using the bundled Solexa image extraction pipeline [version 1.5 or 1.6 (Cassava)]. Sequences (36 bp) were aligned using MAQ software (http://maq.sourceforge.net) using murine genome National Center for Biotechnology Information Build 36 and 37 [University of Cali-

fornia, Santa Cruz
(UCSC) mm8] as the reference genome. Sequences were map- ped using iterative mapping from 36 to 26 bp, excluding reads with more than one mismatch and reads that have a sequence mapping quality Phred score of <50. Reads that had more than
two exact matches were also excluded to correct for sequence bias. As a minimal requirement between 5 and 10 million reads per IP had to be successfully mapped. Analysis of our sequence data were done on the basis of previous models [7,8]. Sequences were extended −400/+600 bp for histone marks representing histone-bound DNA length plus fragment length, or +200 bp for transcription factors, and allocated in 25-bp bins (1.05 °— 108 bins total). Statistically significant enriched bins were identified using a Poissonian background model, generally with
a P value threshold of 10−8 to minimize false-positive results. For enhancer identification, a P value of 10−6 was used for K4me3 to minimize false-negative promoters in our enhancer pool. We used an empirical background model (H3 for histone marks, whole-cell extracts for other factors) that require genomic bins to
be enriched at least threefold above background and extended regions to be enriched at least fivefold above background to correct for nonrandom enrichment observed previously [7]. Enhancer locations were defined by the peak of the enriched regions, and these regions were required to be located at least 1,000 bp away (edge to edge) from a k4me3-enriched region or a known transcriptional start site (TSS) (downloaded from the UCSC table browser; http://genome.ucsc.edu/cgi-bin/hgTables?command=start Refseq), including TSS for miRNAs [7]using
Galaxy (http://main.g2.bx.psu.edu). Enhancer peaks that were within 500 bp of each other were merged into one enhancer. Identification of DNA binding motifs on enhancers. Repeat masked conserved sequences from 200-bp regions under the top enhancer peaks (75 counts or more) were analyzed using Bioprospector [9]
to identify motifs having a width from 6 to 20 nt. Motifs were matched against known transcription factor motifs in TRANSFAC, UNIPROBE (mouse), and JASPAR (core) databases using TOMTOM [10].

**ChIP-Seq cluster and meta-gene analysis.** Enhancers were aligned around the peak loca-

tion, and genes were aligned according to the position of the transcription start site. Around these sites a region of −4 to +4 kbp was selected and subdivided in 200-bp bins. Read density profiles were normalized to the read density per
million total reads per bin around the center of the region. These regions were either left unsorted or clustered using k-means clustering (Euclidean distance) using Cluster 3.0 software (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm) .Heatmaps were generated using Java treeview (http://jtreeview.sourceforge.net/). Meta-gene plots were generated by collapsing all bins and calculating the average read density per bin corrected for the total number of reads

**Correlation of ChIP-seq data.** Correlations were performed on 100-bp bins spanning −4k to +4k bp of each enhancer or TSS site. The maximum value of the region in each sample is defined as the peak. The peaks for all regions from all samples were quantile normalized. The bin values of each region were then rescaled
according to the normalized peak value. Then bin values across all regions from each sample are then concatenated to form a binding vector. Binding vectors from all samples were clustered by hierarchical clustering using Pearson correlation and average linkage (see Hierarchical clustering for more details).

**Preprocessing of gene expression data.** In addition to our our microarray data, a set of 531 microarray data generated using the same platform (GPL4134) was downloaded from the Gene Expression Omnibus database, which will serve as basis for the rowwise z-score transformation described below. Microarray data
were processed and normalized within the array by LOESS and across arrays by quantile normalization using the limma package. To get the standardized expression of each probe values across samples, row-wise z-scores were calculated using all array data per probe across all samples.

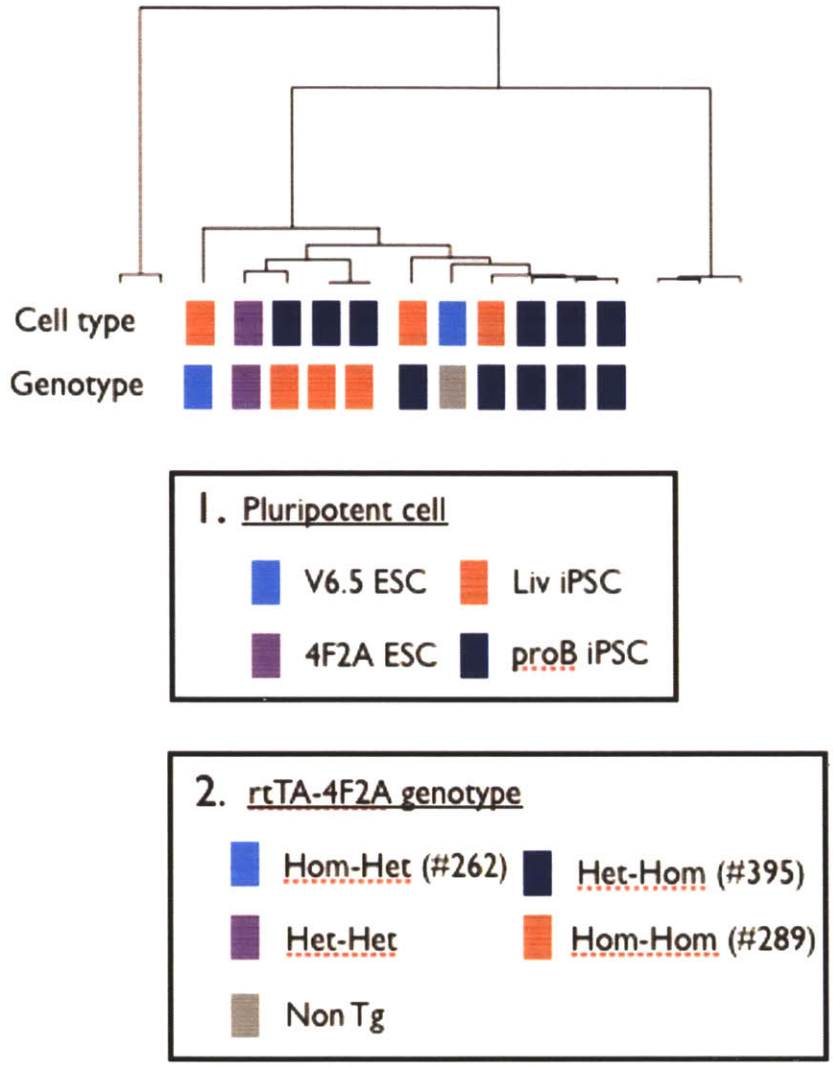
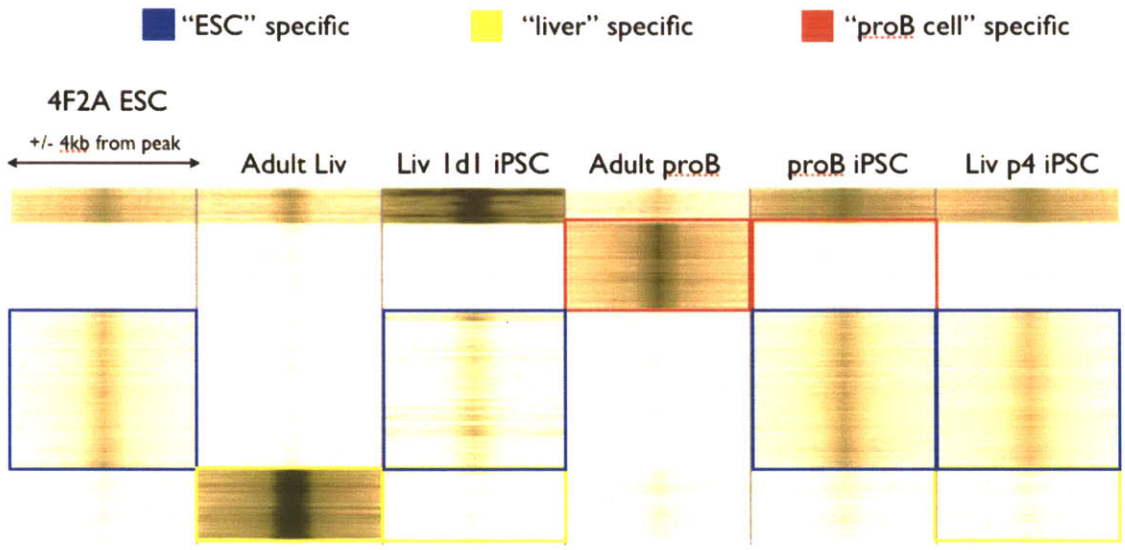



**FIGURE 1:** Genetic background, not tissue-of-origin, correlates with gene expression states between ESCs and iPSCs. Gene expression of microarray data generated in nine pluripotent (3 ESC, 8 iPSC lines) clustered using Cluster3 program. Cell type analysis (#1) shows that liver- or proB-derived iPSCs do not cluseter together. Genetic background, or genotype, analysis shows a better correlation with iPSCs from the same mouse showing higher similarity. Note that ESCs are present within both iPSC-containing groups suggesting iPSCs do not harbor as distinct gene expression state. The far left and right groups respectively, are adult tissues (liver and CD19+ proB cells).

**FIGURE 2**: Global enhancer networks are reset during nuclear reprogramming. Heatmap of 118,935 (only subset shown below) distal enhancers based on H3K4me1 enrichemnt (green) and H3K4me3 depletion in the four indicated cell/tissue types showing tissue-specific enhancer distribution. Four-kilobase pairs around the center of the enhancer region are displayed.



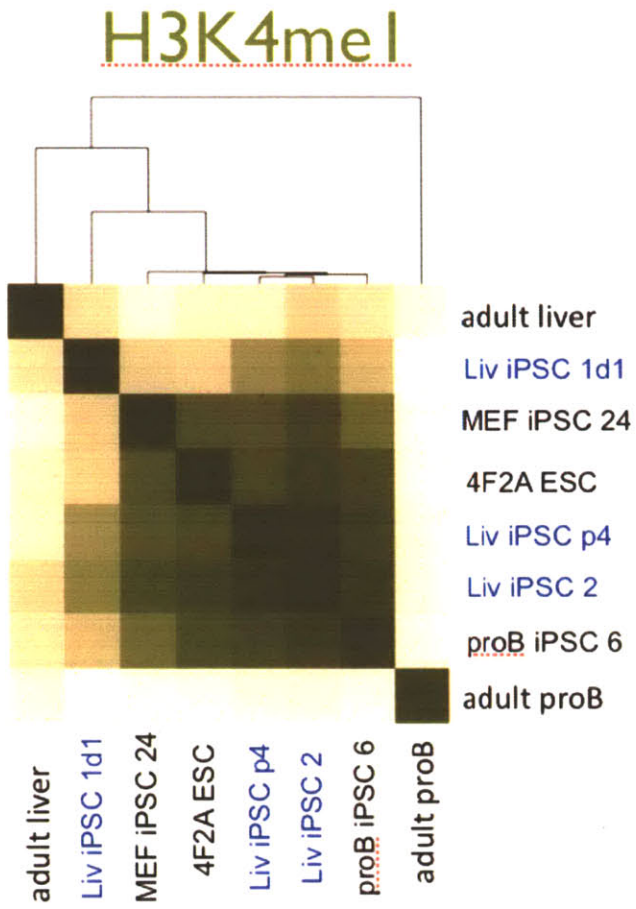



**FIGURE 3:** Majority of iPSC lines reprogram histone modifications associated with cis-regulatory elements. Shown is correlation analysis for the H3K4me1-enriched regions shown in A across the whole set of enhancer regions. Color intensities are a measure of correlation (Peason). Note the closer similarity of Liv 1d1 to Liv p4 and Liv p2. In addition Liv 1d1 is an outlier, not similar to ESCs, by global H3K4me1 pattern and is as different as adult liver compared to ESCs.

**FIGURE 4:** Correlation analysis for the H3K27ac-enriched regions shown in A across the whole set of enhancer regions. Color intensities are a measure of correlation (Peason), which is also indicated by numbers (data not shown). Note the *similarity* of Liv 1d1 to proB6 and V6.5 ESCs, which is in contrast to similar analysis using H3K4me1. NOTE: prior to this analysis H3K27ac data was not divided into two groups i) promoter-associated or ii) distal enhancer-associated therefore the extent of successful reprogramming at enhancer, promoter, or both DNA elements, is unknown.

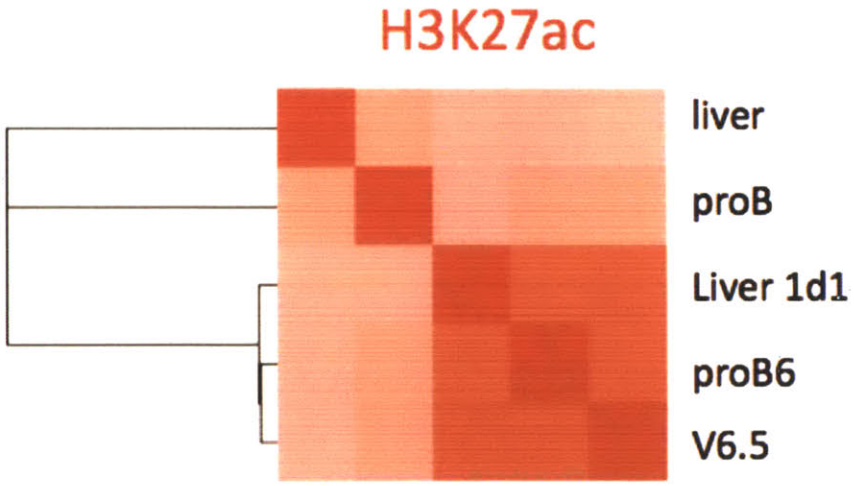


**FIGURE 5:** ChIP-Seq tracks utilizing IGV genome browser to examine enriched regions for specified histone modifications in adult and pluripotent cell types. Liv 1d1 iPSC shows H3K4me1 patterns that reflect the H3K4me3 profile in adult liver, suggesting incomplete reprogramming of this region.

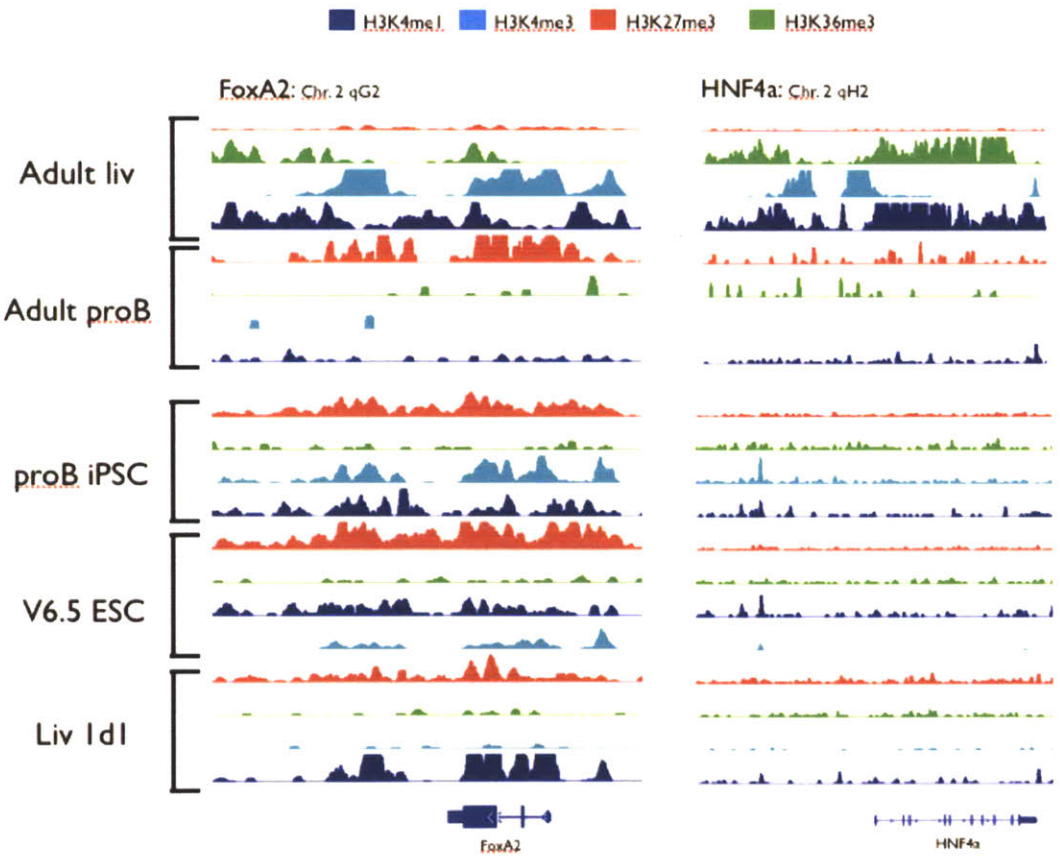


**FIGURE 6**: Incomplete reprogramming of "poised" enhancers from adult somatic cells. H3K4me1 regions show incomplete reprogramming in Liv 1d1 iPSC line and reflect the state of adult liver. Note: Sox4 is not expressed in adult liver.

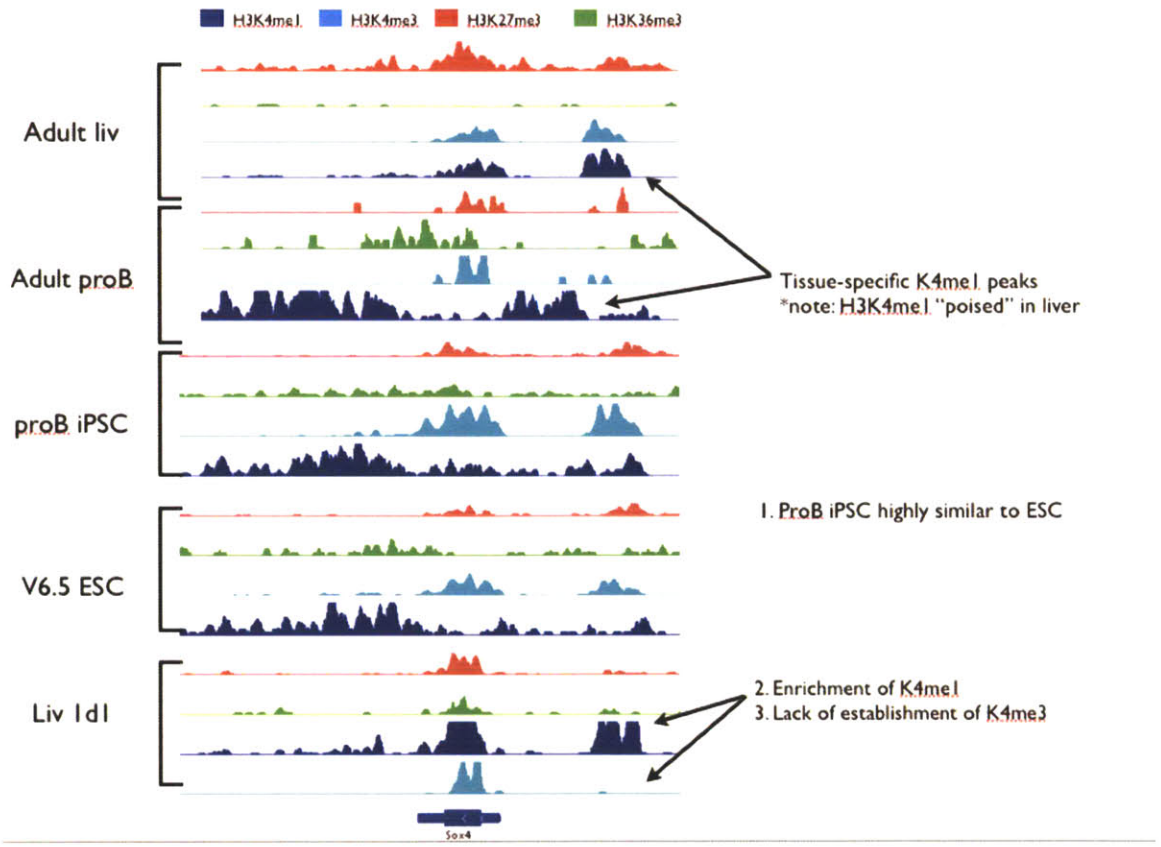
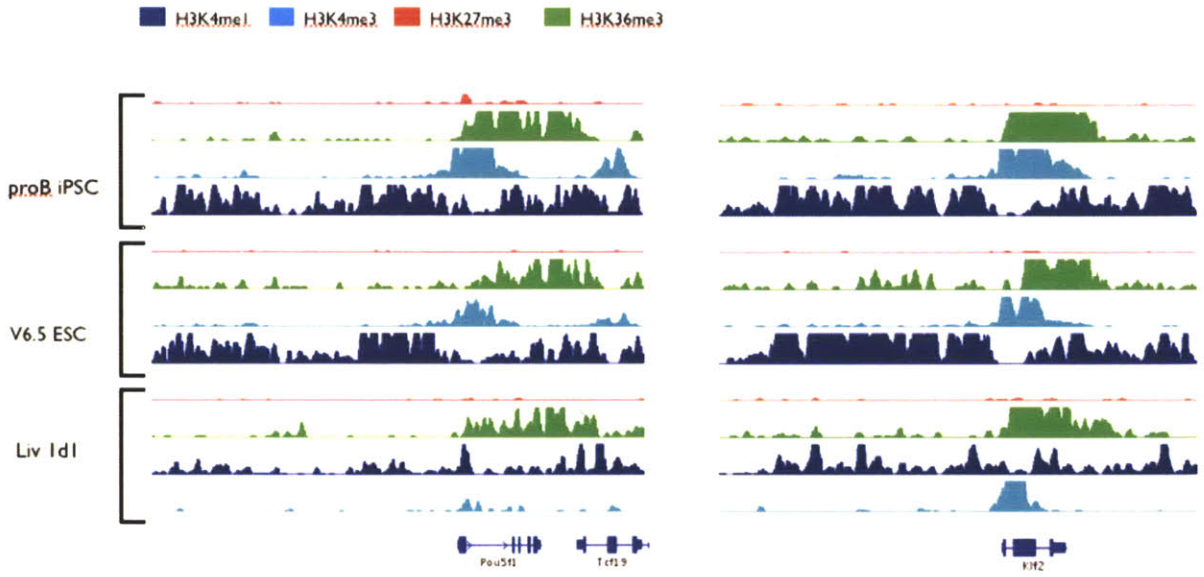


**FIGURE 7**: Incomplete reprogramming of pluripotency genes in Liv 1d1 iPSC line. ChIP-Seq tracks show lack of ESC-like enrichment of H3K4me1 and H3K4me3 in Liv 1d1 iPSC, although these cells express Oct4, and Klf2 RNA as shown by microarray and H3K36me3 signal (below, green).



**REFERENCES CITED:**


1. Kim, K. et al. Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285-290.
2. Polo, J.M. et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* **28**, 848-855.
3. Heintzman, N.D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112 (2009).
4. Heintzman, N.D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318 (2007).
5. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A.* 107(50):21931-6 (2010).
6. Carey, B.W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat Methods* **7**, 56-59.
7. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521-533 (2008).
8. Guenther, M.G. et al. Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev* **22**, 3403-3408 (2008).
9. Liu, X., Brutlag, D.L. & Liu, J.S. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput*, 127-138 (2001).
10. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity between motifs. *Genome Biol* **8**, R24 (2007).